Empirical studies on exchange rate target zones and the microstructure of securities markets

de Jong, F.C.J.M.

Publication date:
1993

Document Version
Publisher's PDF, also known as Version of record

Empirical Studies on Exchange Rate Target Zones and the Microstructure of Securities Markets

Frank de Jong

Empirical Studies on Exchange Rate Target Zones and the Microstructure of Securities Markets

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van dekanen aangewezen commissie in de aula van de Universiteit op donderdag 9 december 1993 te 14.15 uur door

Franciscus Cornelis Johannes Maria de Jong

geboren te Prinsenbeek.

Promotoren: prof. dr. Th.E. Nijman
prof. dr. F. van der Ploeg

MISS PRISM: That would be delightful. Cecily, you will read your Political Economy in my absence. The chapter on the fall of the Indian Rupee you may omit. It's a bit too sensational. Even these metallic problems have their melodramatic side.

Oscar Wilde, The Importance of Being Earnest.

Preface

Some of the chapters of this thesis have appeared elsewhere. Chapter 2 is a slightly revised version of De Jong and Van der Ploeg (1991). Chapter 3 is forthcoming in the *Journal of Applied Econometrics*. Chapter 5 is virtually identical to De Jong, Nijman and Röell (1993).

Acknowledgements

This dissertation is the product of nearly four years work at Tilburg University, the first three years at the Center for Economic Research and the final year in the Department of Econometrics. Both provided a pleasant and inspiring research environment. Several persons deserve special thanks. First of all I'd like to thank my supervisors Rick van der Ploeg and Theo Nijman for their great efforts. I enjoyed cooperating with Roel Beetsma very much. Thanks are also due to Hossein Samiei and Hashem Pesaran for giving me the opportunity to visit Cambridge University. Finally, I thank my parents, brother and sister for their support.

The cover is designed by Maureen de Jong. It shows two French coins from the fifteenth century and a Roman coin from the first century, probably the last successful attempt to create a monetary union in Europe.

Financial support of the Staal Stichting 1966 is gratefully acknowledged.

Contents

Chapter 1

Introduction

This study consists of five empirical essays on European financial markets. As the title suggests, the study is in two parts. Exchange rates in the European monetary system are the subject of the papers in part I. Stock markets are analysed in part II. This chapter serves as an introduction to both parts.

Part I: Exchange Rates

1.1 A brief history of the European Monetary System.

The monetary landscape in the countries of the European Community in the 1980s has been dominated by the participation of many countries in the European Monetary System, the EMS. From 1979 onwards, the EMS has tried to bring exchange rate stability which is seen by many analysts as a necessary prerequisite for economic integration and convergence to a single market within the EC. The exchange rates of currencies that participate in the Exchange Rate Mechanism of the EMS have only a limited fluctuation margin vis à vis the other currencies. In the economic literature, the fluctuation margin is called a **target zone**. The first years of the EMS were not the most successful ones. Especially in the 1982/3 period, when the socialist government in France followed an expansionary economic policy, there were numerous crises, especially involving the French franc, but also the Danish krone and Belgian franc were prone to various devaluations.

After the crisis of March 1983, the French government decided to radically change its economic policy and bring it more in line with the restrictive German fiscal and monetary policy. This started a relatively stable period in the EMS, with few official realignments (except for the Italian lira). New countries like Spain and Portugal entered the EMS and finally in 1990 the British pound was included in the exchange rate mechanism (ERM) of the EMS. The stability lasted until September 1992, when the British pound and the Italian lira dropped out of the ERM, and the Spanish peseta and Portugese escudo were devalued. Only the core currencies of the EMS were unaffected. On August 2, 1993, the EMS all but collapsed under the speculative attacks on the French franc. The EC finance ministers decided to extend the fluctuation margins to ±15%, which is much wider than the ±2.25% used before. All currencies, except the Dutch guilder, are now effectively freely floating against the Deutschemark.

During all the years of the EMS, Germany had the strongest currency, the Deutsche Mark, which has never been devalued against any other EMS currency. German interest rates traditionally are lower than those of other EMS countries. Some economists, e.g. Giavazzi and Giovannini (1989a) argue that the EMS has functioned as a greater Deutschemark area, where the other countries peg their currencies to the Deutschemark. The gain for these countries is increased exchange rate stability and credibility of monetary policy.

1.2 Macroeconomic implications of exchange rate target zones.

An obvious reason for having fixed or managed exchange rates (such as a target zone) is exchange rate stability. Especially in the EC with its common agricultural policy and convergence to one single market, exchange rate stability is of vital importance. Freely floating exchange rates often show large movements which sometimes seem out of line with the underlying economic determinants, cf. the very strong US dollar in the mid 1980s. Mussa (1986) conducted a large cross-country study into (semi-)fixed and flexible exchange rate regimes and concluded that indeed fixed exchange rate regimes increase both nominal and real exchange rate stability. Giavazzi and Giovannini (1989a) conducted a similar study for European exchange rates before and during the EMS and find evidence that the EMS has stabilised exchange rates.

A less obvious reason for countries to join a fixed exchange rate regime is to gain credibility for their monetary policy. Pegging a weak currency to a strong, low-inflation currency such as the Deutsche Mark increases the cost of a high inflation policy, because such a policy leads to forced devaluations, which countries perceive to be costly (Giavazzi and Pagano (1988)). In the absence of exchange rate targets, governments may want to pursue a high inflation policy for several reasons. First, the government might want to create inflation for stabilisation reasons. Barro and Gordon (1983) argue that as soon as nominal contracts are agreed, it is optimal for governments to create surprise inflation to decrease the real cost of labour and thus boost the economy. Rational agents anticipate this policy and increase nominal pay rise demands, up to the point where it becomes too costly for the government to create additional surprise inflation. The economy hence ends up in a 'bad', high inflation equilibrium. Pegging the exchange rate may put the economy back to a 'good', low-inflation equilibrium by increasing the cost of surprise inflation for the government.

Another important reason for inflationary policies is to generate seigniorage, i.e. an inflation tax on nominal money and debt holdings. For some EC countries, mainly in southern Europe, seigniorage traditionally is an important source of government revenues. Participating in the EMS limits the possibility of an inflationary policy and hence diminishes seigniorage revenues. The relation between seigniorage policies and fixed exchange rates is analysed by Van Wijnbergen (1991) and Grilli (1988). Grilli presents a model where the government trades off distortionary taxation with costly inflation, given the desire for exchange rate stability. Grilli demonstrates that entering the EMS shifts the optimal seigniorage level downward and tax revenues upward.

Chapter 2 presents a model for inflation that rests on two building blocks: a tradeoff between costly tax collection and seigniorage, and intertemporal tax/seigniorage smoothing. The government finances its expenditure by tax and seigniorage revenues in an intertemporal setting. Both rules and discretion outcomes are derived. We analyse what entering the EMS changes to the optimal policy rules. It is demonstrated that especially for countries with a high public debt and an inefficient tax collection system, the credibility gains from entering the EMS can be substantial. In the empirical part, several implications of the theoretical model are tested on data from EMS countries and the US. The data provide weak support for one of the theoretical predictions, cointegration between domestic and foreign inflation and tax rates. The model is also confronted with more traditional 'stagflation'-type theories of inflation, which appear to empirically dominate the tax/seigniorage model.

1.3 Interventions and realignments.

In the Exchange Rate Mechanism (ERM) of the EMS, exchange rates are allowed to fluctuate in a band around the central parity. Before August 1993, this band was ±2.25% of the central parity (±6% for certain weak currencies)[1]. Within the given band, the exchange rates are allowed to fluctuate freely. There is a divergence

[1] On August 2, 1993, all bands, except the Deutschemark/Dutch guilder band, were extended to ±15%. This was announced as a temporary measure, but when the old bands will be restored is unknown.

indicator that signals if one of the exchange rates deviates more than 3/4 of its band from a weighted average of all central parities. In practice, this system has not been important, see Giavazzi and Giovannini (1989a, Chapter 2).

More important is the intervention policy if one of the bilateral target zones is under strain. If the exchange rate hits the upper or lower limit of the band, intervention by central banks to support the weak currency is compulsory. There is a Very Short Term Finance facility, under which central banks provide loans to each other to buy the weak currency and sell strong currencies. In principle, the volume of the loans is unlimited, but of course the strong countries often do not want to support a very weak currency under attack indefinitely. If so, there are two options. Either the currency reverts to free floating (as has happened in September 1992 with the pound and the lira) and effectively drops out of the ERM, or the central parity is adjusted, a so called **realignment**. The monetary committee of the EMS has to decide on the time and magnitude of a realignment. Usually, realignments have been small, in the sense that the old and new target zone partly overlap. This is an advantage of a target zone system over completely fixed exchange rates, where the exchange rate always jumps at a realignment. In the EMS, the exchange rate itself needs not jump if the old and new target zone overlap. Nevertheless, here have been several large realignments where the exchange rate made a jump.

1.4 Target zone models of exchange rates.

Since the seminal paper of Krugman (1991), which first circulated in 1988, a large literature on modelling exchange rates in a target zone has emerged. Krugman and Miller (1991) discuss the main characteristics of these models. The approach is dominated by the use of latent variable models in continuous time, which are solved and analysed by using stochastic calculus and various types of 'smooth pasting' conditions. This section gives some background to this literature. It introduces the reader to the models and mathematics used in the target zone literature, and to the major points of interest. This section is not meant as a comprehensive review of the literature. For this, we refer to Svensson (1992) and Bertola (1993).

1.4.1 The Krugman (1991) model.

Krugman (1991) derives his target zone model from an 'asset pricing' equation for the exchange rate

$$(1.1) \qquad e(t) = f(t) + \alpha E de(t)/dt, \qquad \alpha \geq 0,$$

where f(t) is called the economic **fundamental**. For an interpretation of this model, it is instructive to see how it can be derived from the well-known monetary model of exchange rate determination in continuous time[2]:

$$(1.2) \qquad \begin{aligned} m(t) - p(t) &= y(t) - \alpha i(t) \\ m^*(t) - p^*(t) &= y^*(t) - \alpha i^*(t) \\ i(t) - i^*(t) &= E_t de(t)/dt \\ p(t) - p^*(t) &= e(t) \end{aligned}$$

where p is the logarithm of the price level, m the logarithm of the nominal money supply, y the logarithm of real national income, i the instantaneous (very short term) interest rate and e the logarithm of the exchange rate (the price of foreign currency expressed in domestic monetary units); foreign variables are denoted with a *. The first two equations are the domestic and foreign money demand functions, respectively. The third equation is the uncovered interest parity (UIP) condition and the last equation the purchasing power parity (PPP) condition. Solving this model for the exchange rate e, given m and y, one obtains

$$(1.3) \qquad e(t) = (m(t) - m^*(t)) - (y(t) - y^*(t)) + \alpha E_t de(t)/dt,$$

which is precisely (1.1) with $f(t) \equiv (m(t) - m^*(t)) - (y(t) - y^*(t))$. So, in the monetary model, the exchange rate is determined by two factors: the relative money supply (corrected for income differences) and the expected rate of change of the exchange rate itself. Solving (1.3) forward, and assuming that there are no speculative bubbles, one obtains

$$(1.4) \qquad e(t) = \int_0^\infty \exp(-\alpha s) E_t f(t+s) ds.$$

This equation makes clear that the exchange rate is determined by the present discounted value of all expected future fundamentals. The important parameter for the forward-looking behaviour of the exchange rate is the interest semi-elasticity of money demand, α.

---

2 Baillie and McMahon (1989) provide a more elaborate introduction to this model.

The monetary model serves only as an economic motivation of equation (1.1). The target zone literature typically assumes that the fundamental f(t) is an unobserved stochastic process. No attempt is made to observe the fundamental, or to regress it on explanatory variables[3]. The first generation target zone models assumes that f(t) is a *regulated* Brownian motion. The Brownian motion (a continuous time random walk) is described by the following stochastic differential equation

$$(1.5) \qquad df(t) = \mu dt + \sigma dW(t),$$

where W(t) is a Wiener process. The *regulated* Brownian motion follows (1.5) within an interval $(\underline{f},\bar{f})$, but the process is 'regulated' (reflected) at the bounds $\underline{f}$ and $\bar{f}$. Recalling the monetary model, regulation of the fundamental corresponds to intervention in the foreign exchange market by means of expanding or contracting the money supply, in order to keep the exchange rate within the target zone.

The solution of the target zone model has some interesting properties. Most prominent is the non-linear, S-shaped form of the exchange rate within the band. The S-shaped curve and the regulation on the boundaries of the target zone have important repercussions on the stochastic behaviour of exchange rates within a target zone. If the exchange rate were freely floating, it would be a random walk itself, with (by assumption) homoskedastic normal increments. In sharp contrast, the exchange rate in the target zone exhibits mean reversion, non-normal increments and conditional heteroskedasticity. These are well known stylised facts of exchange rates and the target zone model is able to generate these properties endogenously. Beetsma (1991) investigates these properties and concludes that the target zone model is not capable of generating enough non-normality and conditional heteroskedasticity to explain observed exchange rate behaviour.

Other empirical papers that analyse the Krugman model are Smith and Spencer (1991), who use the Method of Simulated Moments to estimate the Krugman target zone model, and Flood, Rose and Mathieson (1990) who test the non-linear relation between exchange rate and fundamental directly. Both papers reject the specification of the Krugman model using EMS data. Bertola and Caballero (1992) and Beetsma and Van der Ploeg (1993) show that the observed probability distribution of EMS exchange rates within the band is not in line with the predictions of the Krugman model. The Krugman model predicts a U-shaped distribution of the exchange rate, with most probability mass concentrated near the edges of the band, whereas

[3] An exception to this rule is the work of Pesaran and Samiei (1992a,b).

empirically one finds a distribution that is concentrated around the central parity.

Chapter 3 is concerned with estimating the Krugman (1991) target zone model and testing its implications for the behaviour of exchange rates in the EMS. Early research used the Method of Simulated Moments to estimate the model. We demonstrate that this estimator is inefficient and computationally unattractive. We propose an efficient Maximum Likelihood estimator that is based on a very accurate and computationally attractive approximation to the exact predictive density of the exchange rate in the continuous time target zone model. Monte Carlo experiments are used to assess the properties of this estimator. In the empirical part we estimate the model with data on recent EMS exchange rates. We find that the Krugman (1991) target zone model is misspecified, because it cannot sufficiently explain the observed kurtosis and conditional heteroskedasticity of relative changes in EMS exchange rates.

1.4.2 Intra-marginal interventions.

Following the empirical rejections, the Krugman model has been extended in several directions, which all add more realism to the model. The most important extensions are the modelling of intra-marginal interventions and realignments.

Although the EMS arrangements only explicitly demand 'marginal' interventions, i.e. interventions if an exchange rate hits the band, central banks often use intra-marginal interventions to manage their exchange rates. For example, Koedijk et al. (1992) present evidence that the Belgian central bank intervened in the foreign exchange market at least once in most of the weeks of the period under consideration (1979 to 1991). Dominguez and Kenen (1992) provide additional evidence on the importance of intra-marginal interventions in the EMS.

Intra-marginal interventions are modelled by Lindberg and Söderlind (1991) as a mean-reverting fundamental instead of the random walk assumed by Krugman. Lindberg and Söderlind estimate their model on Swedish data by the Method of Simulated Moments. The mean reversion is significant, indicating that the Krugman model is misspecified for the Swedish exchange rate. Mean reversion in the fundamentals seems *a priori* important for EMS currencies as well because central banks often use intra-marginal interventions.

1.4.3 Target zone credibility.

Until now, all models that we referred to assumed that the target zone is fully credible. Even without an explicit model for realignments, there are ways to assess the credibility of a target zone. In a series of papers, Lars Svensson demonstrated how to use interest rate differentials to derive the market's expectations of future devaluations. The simplest test of target zone credibility (Svensson, 1991a) is to check whether observed interest rate differentials at several maturities are consistent with the exchange rate staying within the target zone. More complicated is the modelling of realignments. Especially in continuous time, ad hoc assumptions are needed to exclude riskless arbitrage possibilities at a realignment. Bertola and Caballero (1992) assume that there is a fixed probability of realignment if the exchange rate hits the band. If there is a realignment, the exchange rate jumps to its position in the new target zone, but if there is no realignment there is a discrete intervention that puts the exchange rate back to its central parity. Using the Bertola and Caballero model, Bartolini and Bodnar (1991) find a significant realignment risk in the Deutschemark/French franc rate in the 1980s.

Another model for realignments is given by Bertola and Svensson (1993). They assume that there are two state variables: one random walk fundamental and a Poisson jump process for the central parity. In this model, there is a fixed probability per unit of time, independent of the position of the exchange rate within the band, that the central parity is shifted. If one is willing to assume that UIP holds, the interest rate differential is equal to the expected rate of depreciation, which can be decomposed into two parts: the depreciation within the band and the expected rate of devaluation, i.e. parity shifts due to realignments. An estimate of the first part can be obtained from a target zone model, and the interest rate differential is observable, so that one immediately obtains an estimate of the expected rate of devaluation. The second part then gives a measure of time-varying realignment risk.

The two-state model of Bertola and Svensson (1993) with realignment risk is analysed by Rose and Svensson (1991) for the French franc/Deutschemark exchange rate. They use a cubic approximation to the Bertola-Svensson model to predict exchange rate changes within the band and find significant mean reversion within the band. Assuming uncovered interest parity holds, the expected rate of realignment can be assessed from the observed interest rate differentials. Rose and Svensson demonstrate that the EMS bands have never been fully credible, but they also show that it is difficult to predict realignments accurately.

1.4.4 Discrete time target zone models.

A different approach to target zone modelling is taken by Pesaran and Samiei (1992a). They incorporate rational expectations in a discrete-time model with a limited dependent variable, and derive an implicit solution for the expectations variable. Although their model has current expectations instead of future expectations, the model also implies an S-shaped relation between exchange rate and expected fundamental. The relation is however not deterministic but stochastic. The results show that the target zone model fits the data of the Deutschemark/French franc exchange rate during the EMS period better than a model that doesn't take the presence of a band on this exchange rate into account.

Koedijk, Stork and De Vries (1993) also propose to use a discrete time target zone model, but in the spirit of the Krugman model the exchange rate function is forward looking and the fundamental is a random walk with reflecting barriers. Koedijk et al. do not provide an explicit solution to the model, however, but approximate the exchange rate solution by a GARCH process with mean reversion. They show that conditional heteroskedasticity and mean reversion are significant for EMS exchange rates. Moreover, they are able to predict some realignments quite accurately.

In **Chapter 4** a discrete time alternative to the continuous time target zone models is proposed. We account for intra-marginal interventions by assuming that the fundamental is mean reverting. Unlike the assumption made in most continuous time models, the process that generates the fundamental need not be Gaussian. We prove existence of a unique solution and provide a straightforward method to compute the solution. An advantage of the discrete time setup is that, once the discrete time model is solved, the likelihood function and moments of the distribution of the discretely observed data are immediately available. The model is estimated on exchange rates from the EMS. The mean reversion of the fundamental is significant, which suggests that intra-marginal interventions are important for EMS currencies. The deviation from normality is empirically important as well; a Student-t distribution for the innovations in the fundamental is strongly preferred over the normal distribution. Using the estimates of the parameters of the model, time series of expected depreciation within the band are constructed. These series are used to assess the credibility of the EMS target zones.

Part II: Stock Markets

1.5 The quality of securities markets.

In the finance literature, considerable attention has been paid to the quality of securities markets. Important aspects of the quality are transaction costs and liquidity. According to Kyle (1985), liquidity covers such various issues as the cost of a round-trip small transaction *(tightness)*, the ability to absorb large orders (*depth*) and the speed at which prices revert to their efficient price after a transaction (*resiliency*).

Transaction costs and liquidity are important issues for the organisation of a stock market. Several papers raise the question how market structure affects transaction costs and liquidity. Madhavan (1992) and Pagano and Röell (1992) distinguish three classes of trading systems: batch auctions, continuous auctions and dealership markets. In a *batch auction* system all orders are collected periodically (usually one or several times a day). An auctioneer determines an equilibrium price at which buy and sell order flows match, and all transactions are executed at that price. This trading system is most often used for stock series with a relatively small trading volume. For stocks with a high trading volume most exchanges use continuous trading systems. In a *continuous auction* system liquidity is provided by the public limit order book. Market orders are executed immediately at the best available prices in the limit order book. Contrary to the order driven auction system is the quote driven *dealership market*. On such a market one or more designated market makers (dealers) post bid and ask quotes on a bulletin board. Potential traders contact the market makers and usually can negotiate a price for their transaction. Of course, these are highly stylised trading systems and most actual securities markets use hybrid systems. For example, on the New York Stock Exchange the market maker (specialist) competes with the public limit order book in providing liquidity.

Another important difference between stock markets, related to the trading system, is the amount of information about the history of trading that is publicly provided. For example, on most continuous auction markets, the complete limit order book and the prices and quantities of previous trades are publicly available information. On the other hand, on some dealership markets disclosure of trading information is not obligatory, or only with a considerable time lag. These differences in information provision have important consequences for the liquidity of the stock market. Madhavan (1992) demonstrates that if market making is competitive and the trading history is publicly available information, the dealership system and the continuous auction system are equivalent. This result does not hold when we consider a dealership market without publication of trading information. Pagano and Röell (1992) argue that such a market has higher transaction costs (for anonymous traders) than a continuous auction market. On the other hand, traders who can identify themselves as liquidity traders may benefit from the slow disclosure of trading information and directly negotiate a price with a market maker that is better than they would obtain on a continuous auction market.

The aim of our analysis is to provide summary measures of transaction costs and liquidity on the market for French equities. The two main exchanges where French shares are traded are the Paris Bourse and SEAQ International in London. These markets have a different trading structure: the Paris Bourse is best described as a continuous auction market whereas SEAQ is a dealership market. Hopefully, the results shed some light on the advantages and disadvantages of the particular trading systems used by these exchanges.

1.6 The bid-ask spread.

An important implicit transaction cost is the bid-ask spread. In the market micro structure literature (cf. Stoll (1989)), broadly three components of the bid-ask spread are distinguished. The first component can be called processing cost. This includes the pure cost of processing an order and the compensation for the services of a dealer, i.e. the compensation for being in the market and providing liquidity to traders. If processing costs were the only reason for the existence of a bid-ask spread, measurement of the average cost of trading would be easy by averaging observed (quoted) spreads over time. Roll (1984) provides a method to estimate the bid-ask spread from a series of transactions prices if no quote data are available. However, there are two phenomena that complicate this simple picture of transaction costs: inventory control and asymmetric information.

Consider a market where all trades are processed by a market maker or specialist who has an obligation to absorb order imbalances. Because sales and purchases do not come in at the same time, the market maker's inventory usually deviates from the desired level. The market maker runs the risk of price fluctuations on these inventory holdings and if he is risk averse he will demand a compensation for this risk. This intuition is formalised in the model of Ho and

Stoll (1981), who show that the inventory control cost is an increasing function of trade size and share price volatility.

The third type of cost for the market maker arises in the presence of asymmetric information between the market maker and his potential counterparties in trading. This theory was first proposed by "Bagehot" (1971) and formalised in the models of Glosten and Milgrom (1985) and Kyle (1985). A trader with superior private information about the underlying value of the shares will try to buy or sell a large number of shares to reap the profits of this knowledge. This creates an adverse selection problem: the market maker, who is obliged to trade at the quoted prices, incurs a loss on transactions with better informed counterparties. To compensate for this loss he will charge a fee on every transaction, so that expected losses on trades with informed traders are compensated with expected profits on transactions with uninformed "noise" traders. Because the informed parties would tend to trade a large quantity in order to maximise the profits from trading on superior information, the adverse selection cost is related to trade size: large transactions are more likely to be initiated by informed traders than small transactions, cf. Easley and O'Hara (1987) and Madhavan and Smidt (1991). Therefore, the adverse selection cost is an increasing function of trade size, and the market maker's quotes for large transactions will be less favourable than the quotes for small ones. How strong this dependence is determines the depth of the market, i.e. the ability to absorb large orders without moving prices too much.

1.7 Price dynamics and the components of the bid-ask spread.

One of the points to which we pay considerable attention is the composition of the bid-ask spread. In order to estimate the different components of the bid-ask spread, it is necessary to assess the different dynamic effects of adverse selection, inventory control and processing cost. The estimates of the price effects of transactions are interesting in their own right as well. They provide a measure of resiliency, i.e. how long it takes for prices to converge to their new equilibrium price.

An important distinction is between permanent and transitory price effects. Consider the following decomposition of the stock price, due to Hasbrouck (1993):

$$p_t = \mu_t + s_t.$$

The first part, μ_t, reflects the underlying semi-strong form 'efficient price' of the stock. Although equilibrium returns are probably correlated over longer horizons, for the analysis of market micro structure, a good working hypothesis might be that μ_t is a *martingale*, i.e. $E(\mu_t|I_{t-1})=\mu_{t-1}$, where I_{t-1} is the publicly available information at time t-1. It is clear that in the presence of asymmetric information trading affects the expectations of the share's value: a trade initiated by the buyer will cause an upward revision of μ_t. Because μ_t is a martingale, this effect is permanent. The second part, s_t, is a disturbance term that measures the deviation of the transaction prices from the efficient price. These deviations are assumed to be transitory, hence s_t is (weak-sense) *stationary* and in the absence of new shocks the transaction prices eventually converge to the efficient price, μ_t.

The processing cost and inventory control effects belong to the stationary part of the price, s_t. The processing cost is strictly transitory, in the sense that the costs are paid by the initiator of a transaction to the counterparty, but it does not affect future prices at all. The price effect of inventory control will be transitory, because it is expected that in the long run the market maker will rebalance his inventory. If we make the additional assumption that the stationary part is not affected by asymmetric information, the three cost components are identified by their different dynamic effects. Note that this is quite a strong assumption: it implies that all new information revealed by the trade is immediately reflected in the efficient price μ_t.

Summary of Chapter 5.

This chapter analyses the cost of trading French shares on two exchanges, the Paris Bourse and London's SEAQ International. We have available a large data set consisting of all quotes, limit orders and transactions for a two month period. The richness of the dataset allows the analysis of the transaction costs for urgent transactions by the quoted bid-ask spread. In London, market maker quotes are directly available, but in Paris the quoted spread in Paris has to be constructed from the public limit order book. The estimates clearly show that the Paris market is tight, but not very deep. Often the limit order book does not provide enough liquidity for even medium sized transactions. The quoted bid-ask spread may however not give a good indication of actually realised transaction costs. Therefore, we also estimate the realised bid-ask spread. Non-parametric estimates of the realised spread are sensitive to some data imperfections, especially reporting errors in the time of transactions in London. We propose a correction for this timing bias. Moreover, the non-parametric analysis needs an estimate of the 'true' value of the shares. We therefore also propose a model-based estimator based on transaction prices only. The empirical results show that for small to medium-sized transactions the Paris Bourse has lower implicit transaction costs. The market in London, however, is much deeper and provides immediacy for larger trades. Contrary to the predictions of most theoretical models we find that the realised spread is decreasing in trade size, rather than increasing.

Summary of Chapter 6.

In this chapter, we estimate the components of the bid-ask spread and the price effects of transactions on the Paris Bourse. To estimate the components of the spread, we extend the Madhavan and Smidt (1991) model to allow for size-dependent processing costs, but given the lack of inventory data, without inventory control cost. The estimates of this model suggest that the processing cost is lower for large transactions than for small ones. This result explains why sometimes the realised bid-ask spread declines with trades size. In line with theoretical predictions, the adverse selection cost increases with trade size.

In this chapter we also discuss several methods to measure the price effects of trading. Simple non-parametric methods and a more general method based on Vector Auto Regressions are discussed. The empirical results are that the long run price effect of a transaction on the Paris Bourse is much larger than its initial price effect. Moreover, there is no reversal of the trade sign, which suggests that inventory control is unlikely to be relevant on the Paris Bourse. A plausible explanation for the observed effects is that new information is revealed only slowly in the prices. Using the long run price effect as a measure of the adverse selection cost, we estimate that between 50% (for small transactions) and 80% (for large transactions) of the spread is due to asymmetric information.

Part I: Exchange Rates

Chapter 2

Seigniorage, Taxes, Government Debt and the EMS

2.1 Introduction.

Although there has been a considerable literature on the international coordination of stabilisation policies under alternative exchange-rate regimes (e.g., Giavazzi and Giovannini, 1989b; McKibbin and Sachs, 1988; Kenen, 1987; van der Ploeg, 1989) and some literature on the credibility gains associated with pegging one's exchange rate and tying one's monetary policy to that of a well-disciplined foreign central bank (e.g., Giavazzi and Pagano, 1988), the allocative and public-finance aspects of budgetary and monetary policies under alternative exchange-rate regimes have received relatively little attention. The standard argument is that taxes should be smoothed over time and that government debt should be allowed to increase when there is a temporary increase in government spending - due to say a war (Barro, 1979). When seigniorage is also a source of revenues for governments, it is clear that inflation and nominal interest rates should go up and down together with tax rates and that seigniorage and tax revenues should be smoothed because only then the marginal intratemporal and intertemporal distortions of the various types of tax revenues are equalised (e.g., Mankiw, 1987; Grilli, 1989). It can be argued that southern European countries such as Italy have a larger black economy, higher costs of tax collection and thus relatively a greater need for seigniorage revenues than northern European countries such as Germany. Some economists have therefore advocated that there is a crawling peg between the currencies of northern and southern Europe in order to accommodate the required inflation differential (e.g., Dornbusch, 1988). This argument against the European Monetary System is less convincing than it looks at first sight, because it ignores the attempts of countries of the European Monetary System to obtain low inflation and monetary discipline through pegging their currencies to the Deutschemark (Gros, 1988).

The objectives of this chapter are to reconsider the public-finance aspects of budgetary and monetary policies in the light of membership of the European Monetary System and also in the light of the high levels of unemployment faced by

Europe, to analyse the incentives for a surprise inflation tax in order to wipe out government debt, and to provide empirical evidence on these issues for the main European countries. Section 2 examines how the standard theory of tax and seigniorage smoothing should be modified to allow for a desire to have a nominal anchor in the shape of pegging one's exchange rate to the Deutschemark. Section 3 discusses the incentive to levy a surprise inflation tax in order to erode the real value of nominal government debt and considers both discretion and rules outcomes. Section 4 assesses the case for joining the European Monetary System in terms of the monetary discipline the Bundesbank offers to other members. Section 5 departs from the assumption of purchasing power parity and allows for the effects of economic policy on levels of unemployment and considers the implications for optimal budgetary and monetary policies. Section 6 provides empirical evidence on the co-integration of tax rates on the one hand and inflation or nominal interest rates on the other hand and on whether membership of the European Monetary System induces departures from tax/seigniorage smoothing. Section 7 provides some non-nested tests against a more conventional cost-push story of inflation. Section 8 concludes the chapter with a summary of the results.

2.2 Tax/seigniorage smoothing with an EMS-anchor.

Consider the government budget constraint

$$(2.1) \qquad \dot{D} = rD + G - \tau T - \theta mY, \quad D(0) = D_0,$$

where D, G, Y, r, τ, θ and m denote the stock of real government bonds, the level of real exhaustive government spending, the level of real output, the real interest rate, the direct tax rate, the growth rate in the nominal supply of high-powered money and the ratio of money balances to the national income (i.e. the inverse of the velocity of circulation), respectively. It is more convenient to express the items in the government budget constraint as fractions of the national income:

$$(2.1') \qquad \dot{d} = (r-n)d + g - \tau - \theta m, \quad d(0) = d_0,$$

where $d \equiv D/Y$, $g \equiv G/Y$ and $n \equiv \dot{Y}/Y$. The solvency condition requires that the government debt does not grow at a rate faster than the interest rate, that is

$$\lim_{T \Rightarrow \infty} d(T)\exp(-\int_t^T [r(s)-n]ds) = 0.$$

This implies together with equation (2.1′) that the current government debt plus the present value of the stream of future government spending must be less than the present value of the stream of future direct tax and seigniorage revenues:

$$(2.2) \qquad d_0 + \int_0^\infty g(t)\exp[-\int_0^t (r(s)-n)ds]dt \leq \int_0^\infty [\tau(t)+\theta(t)m]\exp[-\int_0^t (r(s)-n)ds]dt.$$

For simplicity, the quantity theory of money is assumed to hold. Hence, m is constant and the rate of inflation, π, equals the excess of monetary growth over the growth in real income, θ-n. Relative purchasing power parity implies $\pi=\pi^*+e$, where π^* denotes the foreign rate of inflation and e denotes the rate of depreciation of the nominal exchange rate. Perfect capital mobility implies that the home interest rate is tied to the foreign interest rate, $r=r^*$.

The government chooses direct taxes and monetary growth to minimise the following welfare-loss function:

$$(2.3) \qquad W \equiv \int_0^\infty \exp[-\int_0^t (r(s)-n)d(s)] \cdot [\tau(t)^2+\beta_1\pi(t)^2+\beta_2 e(t)^2]dt, \quad \beta_1,\beta_2 \geq 0.$$

Hence, the government wishes to minimise the excess burden caused by direct taxes and by inflation. The distortions correspond to the dead-weight losses of taxation and are proportional to real output. There is a slight problem in justifying the cost of inflation in terms of welfare triangles under the money demand schedule, since under the quantity theory this schedule is flat. Nevertheless, even anticipated inflation is costly (e.g., Fischer and Modigliani, 1975; Driffill, Mizon and Ulph, 1989) and it seems sensible to include it in the government's objective function. At the same time the government also wishes to stabilise the nominal exchange rate or, if one does not like the assumption of relative purchasing power parity, to minimise the inflation differential with other countries. The case of floating exchange rates corresponds to $\beta_2=0$ (cf. Mankiw, 1987). In general, the magnitude of β_2 determines the strength of the EMS-anchor (cf. Grilli, 1989).

The first-order conditions for the government are $\tau=\lambda$ and $\beta_1\pi+\beta_2 e=m\lambda$, where λ denotes the marginal cost of an additional unit of government debt. The marginal cost of direct taxes, the marginal cost of inflation (per unit of real money balances) and the marginal cost of government debt should thus be equalised. It follows that in equilibrium

$$(2.4) \qquad \pi = \frac{m\tau + \beta_2 \pi^*}{\beta_1 + \beta_2}.$$

must hold. A regime of floating exchange rates is characterised by $\beta_2 = 0$, so that inflation rates and direct tax rates move up and down together in order to equalise the marginal distortions of both types of revenues. A regime of fixed exchange rates is characterised by $\beta_2 \Rightarrow \infty$, so that home inflation is tied to the foreign inflation rate. In general, this theory predicts that EMS countries without capital controls have a convergence of inflation rates (witness, for example, the closeness of inflation rates in the Netherlands and Germany) whilst for other non-EMS countries inflation rates (or nominal interest rates) and direct tax rates should be co-integrated and more or less independent of the foreign inflation rate. Of course, some EMS countries (such as Italy and France) had capital controls in which case they should have been able to follow an independent monetary policy and engage in some tax/seigniorage smoothing without completely forsaking the objective of relatively fixed exchange rates.

The first-order conditions also require that direct tax rates follow a random walk, $\dot{\tau}^e(s,t) = 0$, $s > t$, where $\dot{\tau}^e(s,t)$ denotes the expectation of $\dot{\tau}(s)$ given all information available at time t. The reason is that the government smooths the distortions of raising tax revenues over time. Similarly, $\dot{\pi}^e(s,t) = \beta_2/(\beta_1 + \beta_2)$ $\dot{\pi}^{*e}(s,t)$ for $s > t$, must hold.

Given equation (2.4) and expectations about future government spending, the government must choose the level of taxes and of seigniorage revenues to satisfy the intertemporal budget constraint (2.2). For a constant real interest rate, one obtains:

$$(2.5a) \qquad \tau = \frac{(r-n)d - nm + g_p - m'\pi^*_p}{1 + m^2/(\beta_1 + \beta_2)},$$

$$(2.5b) \qquad \pi = \theta - n = \frac{[(r-n)d - nm + g_p]m + \beta_2[\pi^* + m^2(\pi^* - \pi^*_p)/(\beta_1 + \beta_2)]}{\beta_1 + \beta_2 + m^2}$$

where $m' \equiv \beta_2 m/(\beta_1+\beta_2)$ and the permanent level of exhaustive government spending is defined as

$$(2.6) \qquad g_p(t) \equiv (r-n) \int_t^{\infty} \exp[-(r-n)(s-t)]g^e(s,t)ds.$$

The permanent level of the foreign inflation rate, π_p^*, is defined in a similar fashion. A permanent increase in the level of exhaustive government spending, caused by expected increases in future government spending, requires higher taxes, seigniorage revenues and inflation, but a temporary increase is financed by issuing government debt. An increase in the expected level of future foreign inflation raises anticipated seigniorage revenues and thus permits a cut in current direct taxes, seigniorage revenues and inflation. A temporary increase in foreign inflation raises inflation, but leaves the optimal direct tax rate unaffected. A permanent increase in foreign inflation leads to a smaller increase in inflation and to a cut in the direct tax rate. The public sector borrowing requirement can be written as

$$(2.7) \qquad \dot{d} = g - g_p - \frac{\beta_2 m}{\beta_1+\beta_2}(\pi^*-\pi_p^*),$$

so that, when times are bad in the sense that exhaustive government spending exceeds its permanent level or foreign inflation is (and thus seigniorage revenues are) below its permanent level, the government borrows.

The above theory of tax/seigniorage smoothing corresponds to an extension of the familiar "Golden Rule of Finance", which says that governments should tax in order to finance government consumption and should borrow in order to finance investment projects with a market rate of return. The reason is that investment projects with a market rate of return leave the permanent level of government spending (net of the return on investment projects) unaffected, whilst government consumption increases the permanent level of government spending.

Some would argue that the desired inflation rate should equal minus the real interest rate as full liquidity requires driving the nominal interest rate, i, to zero (cf. Friedman, 1969). In that case, the term $\beta_1\pi^2$ in the welfare-loss function (2.3) should be replaced by $\beta_1(\pi+r)^2$ and one finds that:

$$(2.8a)\qquad i = r + \pi = \frac{m\tau + \beta_2 i^*}{\beta_1 + \beta_2} = \frac{[(r-n)(d+m)+g_p]m \ + \ \beta_2[i^* + m^2(i^* - i_p^*)/(\beta_1+\beta_2)]}{\beta_1 + \beta_2 + m^2},$$

$$(2.8b)\qquad \tau = \frac{(r-n)(d+m) \ + \ g_p \ - \ m' i_p^*}{1 \ + \ m^2/(\beta_1 + \beta_2)},$$

where again $m' = \beta_2 m/(\beta_1 + \beta_2)$. Only when there are no distortions from raising direct tax revenues and when there is no EMS-anchor ($\beta_1 \Rightarrow \infty$) are nominal interest rates driven to zero and is full liquidity attained. In general, when there is a need to raise revenues for the public sector, there is a trade-off between zero tax distortions and the desired (negative) level of inflation which, typically, leads to positive optimal tax rates and positive optimal inflation rates (cf. Phelps, 1973).

When there is a wide-spread black economy and the dead-weight burden of non-monetary taxes is relatively much higher than that of inflation (β_1, low), it is optimal to extract relatively little revenues from conventional taxation and relatively a lot from seigniorage. It follows that for such economies inflation and nominal interest rates are rather higher than for economies with a relatively insignificant black economy. One could argue that this is a reason why some southern European countries find it optimal to have higher inflation rates than northern European countries (cf. Dornbusch, 1988, Canzoneri and Rogers, 1990).

Under a regime of floating exchange rates nominal interest rates are now co-integrated with direct tax rates, whilst under fixed exchange rates they are tied to foreign nominal interest rates. The public sector borrowing requirement can now be written as

$$(2.9)\qquad \dot{d} = g - g_p - \frac{\beta_2 m}{\beta_1 + \beta_2}(i^* - i_p^*),$$

so that when foreign nominal interest rates are high and expected to fall the government pays off its debt.

2.3 Nominal bonds and the surprise inflation tax: Rules versus discretion.

So far, we have assumed that the government issues real (or indexed) bonds. In practice, governments usually issue bonds with a guaranteed nominal rate of return, $i \equiv r+\pi^e$, where π^e denotes the expected rate of inflation and r denotes the ex-ante real interest rate. This implies an appeal to the Fisherian hypothesis, so that the (unobservable) ex-ante real interest rate (or real interest rate for short) is determined by consumer preferences and production technologies, more or less independent of the expected inflation rate. For given tastes and technologies, any change in the nominal interest rate must then be due to a change in the expected inflation rate. The realised or ex-post real interest rate, $i-\pi$, is relevant for the borrowing and lending activities of the government. It follows that the government budget constraint becomes

$$\text{(2.10)} \qquad \dot{d} = (r+\pi^e-\pi)d + g - \tau - \theta m,$$

where d(0) is a non-predetermined variable as the price level at time zero is non-predetermined. The realised or ex-post interest rate decreases with unanticipated inflation, which is one way in which the government can reduce the level of the inflation-corrected deficit and reduce the growth of the debt-GDP ratio[1].

Two outcomes should be considered (cf. Kydland and Prescott, 1977). The first is the **rules** outcome, which prevails when the government is believed to have sufficient discipline not to succumb to a surprise inflation tax and wipe out the real value of outstanding debt. The government can then credibly influence expectations of the private sector, so that in the determination of its optimal policy it can assume that $\theta^e=\theta$ and $\pi^e=\pi$. Hence, the rules outcome (denoted by the superscript R) is observationally equivalent to a situation where the government issues real or indexed bonds. However, the rules must be enforced or else the government has an incentive to renege and reoptimise. This time-inconsistent behaviour usually takes the form of a surprise increase in monetary growth, because this permits a corresponding reduction in distortionary taxes.

[1] The government debt-GDP ratio, d, is no longer a predetermined variable, because an increase in the level rather than the growth of the nominal supply of money induces an equal increase in the price level and thus wipes out the real value of government debt at "the stroke of a pen". However, the present analysis is concerned with unanticipated increases in inflation.

The second is the **discretion** outcome (denoted by the superscript D), which occurs when the government cannot make credible announcements about monetary policy and must therefore take θ^e as given. It follows that equation (2.4) is replaced by

$$(2.11) \qquad \pi^D = \frac{(m+d)\tau^D+\beta_2\pi^*}{\beta_1+\beta_2},$$

so that for a given tax rate inflation is higher (lower) in equilibrium than under the rules outcome when the government has accumulated debt (assets). The reason is that the private sector does not believe the announcement of lower monetary growth, because the government is then tempted to levy a surprise inflation tax. Upon substitution into the intertemporal government budget constraint, one obtains for $d_p>0$:

$$(2.12a) \qquad \tau^D = \left(\frac{(r-n)d-nm+g_p - m'\pi_p^*}{1 + m(m+d_p)/(\beta_1+\beta_2)}\right) < \tau^R,$$

$$(2.12b) \qquad \pi^D = \left(\frac{[(r-n)d-nm+g_p](m+d)}{\beta_1+\beta_2+m(m+d_p)}\right) + \beta_2\left(\frac{\pi^* + m''[(m+d)(\pi^*-\pi_p^*)+(d_p-d)\pi^*]}{\beta_1+\beta_2+m(m+d_p)}\right),$$

where again $m'=\beta_2 m/(\beta_1+\beta_2)$ and $m''=m/(\beta_1+\beta_2)$. It is easy to demonstrate that, when $g=g_p$, $\pi=\pi_p^*$ and $d=d_p>0$, $\tau^D<\tau^R$ and $\pi^D>\pi^R$ must hold. Hence, a lack of monetary discipline for countries with a large government debt leads to lower taxes and higher inflation than would be the case when governments enjoy a reputation for sticking to rules.

2.4 The case for joining the EMS.

One could argue that the fact that the Banca d'Italia is dependent on the whims of politicians is the reason why Italy with a government debt of around 100% of its national income has a higher inflation rate and a greater reliance on seigniorage revenues than Germany with the excellent monetary discipline afforded by the Bundesbank and lower government debt. Another reason might be that Italy has a greater black economy and a less efficient tax system than Germany (proxied by relatively low values of β_1 and β_2)[2] and therefore finds it optimal to extract relatively more from seigniorage than from direct taxation (cf. Canzoneri and Rogers, 1990). Indeed, some have argued that these are good arguments why there should be a crawling peg between the currencies of northern and southern Europe as this would accommodate the required difference in inflation rates (e.g., Dornbusch, 1988).

However, these two arguments for Italy against the EMS compare the discretion outcome for Italy with the rules outcome for Germany and completely ignore the monetary discipline the Bundesbank gives to the Banca d'Italia under a firm membership of the EMS. Some consider the gain in central bank credibility and the accompanied tying of one's hands as the main advantage of the EMS[3]. In that case, the appropriate question to ask is whether Italy is better off under a discretionary regime with floating exchange rates as would be the case under an EMS with wide bands (small β_2) or better off under a rules outcome with fixed exchange rates as would be the case under an EMS with narrow bands (large β_2) or possible under an EMU. The point is that under a float Italy has to stand on its own feet and make do with the reputation of a central bank which is not independent of the wishes of spending ministers and trade unions, but under the EMS Italy can enjoy the advantages in terms of monetary discipline of the Bundesbank. Of course, the price Italy may pay for membership of the EMS is a sub-optimal government revenue mix.

One should then trade off the gains in monetary discipline derived from tying monetary policy to the policy of the Bundesbank against the losses arising from a sub-optimal government revenue mix, i.e. too low seigniorage revenues and too high distortionary tax rates (cf. Gros, 1988). For example, comparison of a discretion outcome under a pure float (β_2=0, denoted by the superscript F) with a rules

---

[2] Indeed, Barro's (1979) argument to include taxes in the government's welfare-loss function was to capture the government's costs of collecting taxes. The idea here is that these costs are for Italy much higher than for Germany. Equations (2.5)-(2.6) or (2.9)-(2.10) then immediately show that ceteris paribus Italian tax rates are lower than German tax rates whilst Italian inflation is higher than German inflation.

[3] Giavazzi and Pagano (1988) discuss this within a context where central banks and unions are locked into nominal wage contracts, and where the former have an incentive to engage in surprise monetary expansions in order to erode the real value of wages and boost employment. Wage indexation used to be prevalent throughout Europe, so that there presumably is not much of a possibility to wipe out the real value of predetermined nominal wage contracts with unanticipated inflation. This is why it is reasonable to consider the incentives for wiping out the real value of government debt. The interaction between both effects are considered in van der Ploeg (1991).

outcome under the EMS ($\beta_2 \Rightarrow \infty$, denoted by the superscript E) yields when Germany has zero inflation:

$$\text{(2.13a)} \qquad \pi^F = \left(\frac{[(r-n)d-nm+g_p](m+d)}{\beta_1+m(m+d_p)}\right) > \pi^E = \pi^* = 0,$$

$$\text{(2.13b)} \qquad \tau^F = \left(\frac{(r-n)d-nm+g_p}{1+m(m+d_p)/\beta_1}\right) < \tau^E = (r-n)d-nm+g_p.$$

It can be shown (for the case of $g=g_p$) that it pays countries to join the EMS as long as $(\beta_1-m^2)d>(\beta_1+m^2)m$. Hence, countries with a large priority to keeping inflation low (high β_1) and a large outstanding nominal government debt have an incentive to join the EMS, because the losses arising from the increase in tax distortions are outweighed by the gains arising from the increase in monetary discipline. However, countries with a large nominal government debt and which care more about removing tax distortions than about achieving low inflation are less keen to join the EMS. This may be a reason why the United Kingdom was less keen to join the EMS than Italy. Germany achieves a stable price level, because this is the sole objective of the Bundesbank and because the Bundesbank is independent of the fiscal authorities in Germany.

This section considered the implications of the use of a surprise inflation tax to wipe out the real value of outstanding nominal government debt. Similar arguments can be made (e.g., Calvo, 1978; Turnovsky and Brock, 1980; Barro, 1983) when the asset menu includes no (or only real) government bonds and the demand for money depends negatively on the expected rate of inflation (e.g., $m=\phi(r+\pi^e)$, $\phi'<0$). The rules outcome then also typically leads to lower inflation in equilibrium than the discretion outcome. If one allows for the ongoing strategic interactions between the monetary authorities and private agents, the government finds it optimal to have temporary bouts of inflation in order to accumulate government assets. The resulting interest income finances the increase in government spending (Obstfeld, 1991; van der Ploeg, 1991).

2.5 Public finance aspects of unemployment policy.

The basic model of tax and seigniorage smoothing has been extended to allow for an EMS-anchor (Section 2.2) and for nominal bonds, expectations and surprise inflation taxes (Sections 2.3 and 2.4). Here the basic model is extended to allow for the endogenous determination of government spending. It also departs from purchasing power parity by having imperfect substitution between home and foreign goods, and allows for the use of budgetary policies to fight unemployment. Without such modifications the model only seems relevant for long-run issues.

Aggregate demand increases when government spending or world trade increases, when the real exchange rate depreciates, when taxes or the rate of interest falls, and when real money balances (wealth) increase. It is therefore given by

$$\text{(2.14)} \qquad y^d = \bar{\alpha}_0 + \bar{\alpha}_1 w^* + \bar{\alpha}_2 g + \bar{\alpha}_3 m - \bar{\alpha}_4 \tau - \bar{\alpha}_5 r^* + \bar{\alpha}_6 c, \quad \bar{\alpha}_i \geq 0,$$

where y^d and w^* denote the ratios of the demand for home goods and world trade to the full-employment level of the national income, Y, and c denotes the real exchange rate (the relative price of foreign goods in terms of home goods). Aggregate supply increases with the capital stock and decreases with the real producers' wage. Since the wedge between the real producers' wage and the real consumers' wage increases when the tax wedge increases or the real exchange rate depreciates, aggregate supply can be written as

$$\text{(2.15)} \qquad y^s = \bar{\alpha}_7 k - \bar{\alpha}_8 c - \bar{\alpha}_9 \tau - \bar{\alpha}_{10} \omega + \bar{\alpha}_{11},$$

where y^s and k denote the ratios of the supply of home goods and the capital stock to the full-employment level of national income and ω denotes the real consumers' wage. A depreciation of the real exchange rate increases import and consumers' prices and thus increases, in exactly the same way as an increase in taxes does, the wedge between producers' and consumers' wages and reduces the aggregate supply of goods. For simplicity, the beneficial effects of unanticipated inflation on aggregate supply are ignored. This can be justified when all wage contracts are indexed to the price level or when one restricts attention to rules outcomes. Equilibrium in the goods market ($y^d = y^s \equiv y$) yields an expression for the real exchange rate, which can then be used to solve for the unemployment rate:

$$\text{(2.16)} \qquad u \equiv 1-y = \alpha_0 - \alpha_1 w^* - \alpha_2 g - \alpha_3 m + \alpha_4 \tau + \alpha_5 r^* - \alpha_7 k + \alpha_{10} \omega,$$

where $\alpha_i \equiv \bar{\alpha}_i/(\bar{\alpha}_6 + \bar{\alpha}_8) \geq 0$, $i \neq 4$ and $\alpha_4 \equiv (\bar{\alpha}_4 + \bar{\alpha}_9)/(\bar{\alpha}_6 + \bar{\alpha}_8) \geq 0$. In contrast to the familiar Mundell-Fleming model with nominal wage rigidity, a fiscal expansion has real effects because it induces an appreciation of the real exchange rate, a cut in the wedge and thus a drop in unemployment. The labour market fails to clear quickly, since the real consumers' wage is assumed to be fixed by institutions at a too high level. Alternatively, in a bargaining model the level of the real consumers' wage decreases with the unemployment rate in which case the α_i in (2.16) would become smaller in magnitude.

The government chooses τ, θ and g to minimise the following welfare loss function:

$$\text{(2.17)} \qquad W \equiv \int_0^\infty \exp[-(r-n)t][\tau^2+\beta_1\pi^2+\beta_2 e^2+\beta_3 u^2+\beta_4(\bar{\bar{g}}-g)^2]dt,$$

where $\bar{\bar{g}}$ denotes the desired ratio of government spending to the national income. The government is now also concerned about achieving full employment and a decent size of the public sector. These additional concerns seem to be a very important feature of the political economy of western Europe. The political colour of the incumbent government matters: left-wing governments tend to have a higher value of $\bar{\bar{g}}$ than right-wing governments. The first-order conditions yield

$$\text{(2.18)} \qquad \tau + \beta_3\alpha_4 u = \beta_3\alpha_2 u + \beta_4(\bar{\bar{g}}-g) = [\beta_1\pi + \beta_2(\pi-\pi^*)]/m = \lambda,$$

and $\dot{\lambda}^e=0$, where λ denotes the marginal cost of an additional unit of government debt. This gives rise to two co-integrating relationships between g, τ and π:

$$\text{(2.19a)} \qquad \pi = \frac{m\tau+\beta_2\pi^*+\beta_3\alpha_4 m\cdot u}{\beta_1+\beta_2},$$

$$\text{(2.19b)} \qquad \tau = \beta_3(\alpha_2-\alpha_4)u + \beta_4(\bar{\bar{g}}-g).$$

An increase in unemployment, due to, say, a world recession or an increase in the real consumers' wage leads to an increase in inflation and seigniorage revenues or a cut in taxes (given π^*). It also leads to higher government spending or higher direct tax rates. There is (given $\bar{\bar{g}}$, u and π^*) a positive association between inflation and tax rates and a negative association between inflation or tax rates and government spending. Upon substitution into the intertemporal government budget constraint, one obtains:

$$\text{(2.20a)} \qquad \tau = -\beta_3\alpha_4 u + \left[\frac{(r-n)d-nm+\bar{\bar{g}}_p-m'\pi^*_p+\bar{\beta}_3u_p}{1+m^2/(\beta_1+\beta_2)+(1/\beta_4)}\right],$$

$$\text{(2.20b)} \qquad g = \bar{\bar{g}} + (\beta_3\alpha_2/\beta_4)u - \left[\frac{(r-n)d-nm+\bar{\bar{g}}_p-m'\pi^*_p+\bar{\beta}_3u_p}{\beta_4[1+m^2/(\beta_1+\beta_2)]+1}\right],$$

$$\text{(2.20c)} \qquad \pi = \left[\frac{(r-n)d-nm+\bar{\bar{g}}_p+\bar{\beta}_3u_p-m'\pi^*_p}{(\beta_1+\beta_2)[1+(1/\beta_4)]+m^2}\right]m + \left[\frac{\beta_2}{\beta_1+\beta_2}\right]\pi^*,$$

$$\text{(2.20d)} \qquad \dot{d} = \bar{\bar{g}}-\bar{\bar{g}}_p - m'(\pi^*-\pi^*_p) + \bar{\beta}_3(u-u_p),$$

where $m'=\beta_2 m/(\beta_1+\beta_2)$ and $\bar{\beta}_3 \equiv \beta_3[(\alpha_2/\beta_4)+\alpha_4]>0$. The signs of these and other budgetary and monetary responses are presented in Table 2.1.

Table 2.1 Optimal budgetary and monetary responses.

	$\bar{\bar{g}}$	$\bar{\bar{g}}_p$	u	u_p	π^*	π^*_p	$(r-n)d$	n
τ	0	+	-	+	0	-	+	-
g	1	-	+	-	0	+	-	+
$g-\tau$	1	-	+/-	-	0	+	-	+
$(r-n)d+g-\tau$	1	-	+/-	1	0	+	+	+
π	0	+	0	+	+	-	+	-
$\dot{d}$	1	-1	+	-	-	+	0	0

A transitory increase in the desired level of exhaustive government spending leads to an equal increase in the actual level of exhaustive government spending, which is financed entirely by borrowing. A permanent increase in the desired level of exhaustive government spending leads to a smaller increase in actual exhaustive government spending and to an increase in direct taxes, inflation and seigniorage revenues, but does not affect government borrowing. An expected increase in the desired level of future exhaustive government spending implies a cut in current exhaustive government spending, an increase in current tax and seigniorage revenues, and current saving of the government. A high level of debt service does not affect government borrowing, but leads to high taxes and monetary growth and to low levels of exhaustive government spending.

A temporary increase in unemployment leads to government borrowing, a cut in taxes and a boost in exhaustive government spending. Since a large component of government spending consists of unemployment benefits and other transfers, it is not surprising that current government spending should respond to current unemployment. A permanent increase in unemployment, however, leaves government borrowing unaffected and gives rise to an increase in both seigniorage revenues and inflation. The effect on taxes and exhaustive government spending is ambiguous, because both instruments stimulate unemployment and both are costly. If government spending is more (less) effective than taxes to cut unemployment, i.e. $\alpha_2>\alpha_4$ ($\alpha_2<\alpha_4$), then government spending should be increased (taxes should be cut).

A temporary increase in foreign inflation leads to a smaller increase in inflation and seigniorage revenues, which is fully used to pay off government debt. A permanent increase in foreign inflation leaves government borrowing unaffected and leads to a smaller increase in inflation. It is also used to cut taxes and boost exhaustive government spending. An increase in real growth allows a cut in tax and inflation rates and an increase in government spending.

The present model of tax/seigniorage smoothing with an EMS-anchor has been extended to allow for a number of important features such as unemployment, imperfect substitution between home and foreign goods, and endogenous government spending. Of course, it is possible to include other features which may be of interest. For example, one could allow for capital controls, which in the past have allowed countries such as Italy and France to have fixed, but adjustable exchange rates **and** an independent monetary policy (witness the differential between on-shore and off-shore interest rates). Since most of the capital markets of Europe have been liberalised from 1st July 1990, this extension will become less relevant.

2.6 Empirical evidence on the co-integration of inflation and tax rates.

In this section we present some empirical results on the model of tax and seigniorage smoothing presented in Sections 2.2, 2.3 and 2.5. The analysis is undertaken for the US and for all EEC countries with the exception of Luxembourg which has no independent monetary policy. The aim of the empirical analysis is twofold. First, from a theoretical point of view we want to investigate the causes of inflation in Europe and judge which model seems best fit to explain the historic pattern of inflation. Second, we look for cross-country differences in inflation experience and public-finance policies. The results may give some indication of the steps that are to be undertaken for the "convergence" of monetary and fiscal policy that may be needed for a monetary union in Europe.

Various authors have investigated the intertemporal theory of tax and inflation smoothing for the United States. Mankiw (1987) finds a positive correlation, both in levels and in first differences, between nominal interest rates and federal tax revenues. Mankiw takes this as evidence supporting the theory. Trehan and Walsh (1990) note that tax and inflation smoothing imply that both tax and inflation rates must have a unit root. Hence, equalising the intratemporal distortions of tax and inflation rates leads to a co-integrating relation between tax and inflation rates. The presence of a long-run or co-integrating relation can be tested by testing for a unit root in the residuals of a regression in levels, see Engle and Granger (1987). The rejection of a unit root in the residuals points at the existence of such a long run relation, i.e. there is co-integration between tax and inflation rates. Trehan and Walsh (1990) cannot reject the null hypothesis of no co-integration between US tax and inflation rates, not even when they allow for nonstationarity in the velocity of circulation (m^{-1}). Moreover, the theory of tax and seigniorage smoothing can explain only a very small part of the observed variation in inflation. Trehan and Walsh conclude that revenue smoothing considerations have not been significant elements in determining the behaviour of seigniorage in the US.

Poterba and Rotemberg (1990) find a positive relation between after-war tax rates and inflation rates in the US, but this result does not carry over to the four other large countries they study, namely, France, Germany, Japan and the UK. Grilli (1988) reports evidence of co-integration between tax and inflation rates for the EEC countries (excluding Portugal and Luxemburg). He also extends the model with an exchange rate anchor to allow for the European Monetary System, in a way similar to ours. Grilli reports evidence of co-integration between domestic inflation and tax rates and foreign inflation for some countries. He also finds that the effect of foreign inflation seems to be stronger for the large countries, Germany and the UK especially, which seems to contradict the hypothesis that it is particularly small countries who tie their exchange rate to the Deutschemark and are willing to give up an independent monetary policy.

Much of previous empirical work has relied on testing for unit roots in tax and inflation series, and on testing for co-integration between those variables. In our view there are some problems associated with this type of research. First, it relies heavily on statistical tests for unit roots, which have a notoriously low power in small samples (e.g. Christiano and Eichenbaum, 1989). Second, the equations often do not allow for any dynamics and give only estimates of long-run (co-integrating) parameters or, if the model is estimated in first differences, of short-run effects. The omission of dynamics in the equation results in a loss of efficiency in estimating the parameters (see Wickens and Breusch, 1988), especially in small samples like ours with only 25 observations. Domowitz and Hakkio (1990) argue that in the presence of costs of adjustment the resulting dynamic pattern can be expressed as an error-correction mechanism. For the problem under consideration, it is very plausible that there are significant costs in adjusting tax rates.

This is why we prefer to estimate the relation between inflation, tax rates and foreign inflation by means of an error-correction mechanism of the form:

$$\Delta\pi_t = \gamma_0 + \gamma_1\Delta\tau_t + \gamma_2\Delta\pi^*_t + \gamma_3(\pi_{t-1} - \delta_1\tau_{t-1} - \delta_2\pi^*_{t-1}) + \varepsilon_t. \tag{2.21}$$

If necessary, lagged first difference terms could be added, but this specification appeared to be sufficient to guarantee serially uncorrelated disturbances in our applications. The parameters γ_1 and γ_2 are the initial effects on inflation of a rise in taxes or foreign inflation. Trehan and Walsh (1990) argue that one should not pay too much attention to such short-run effects, because the tax-inflation tradeoff can in the short run be obscured by a variety of other factors. Some of these issues are investigated in Section 2.7, where we present an alternative model of inflation. The more important parameters are δ_1 and δ_2. These parameters show the magnitude of the long-run effect on inflation of a rise in tax rates and foreign inflation.

If one is willing to accept that π_t, τ_t and π^*_t are co-integrated, then $(-1,\delta_1,\delta_2)$ is the co-integrating vector of these variables. The significance of the long-run relations can be assessed in two ways. First, the error-correcting effect should be significant, i.e. γ_3 should be negative and significantly different from zero. Note that if the time series of π_t, τ_t and π^*_t contain unit roots, the error-correction term is non-stationary under the null-hypothesis $\gamma_3=0$ and consequently the usual critical values for the t-statistics do not apply. Pagan and Wickens (1989) mention that in such cases, given the significance level, the critical values are higher.

This brings us to the next point, estimation of the model. One cannot estimate equation (2.21) by OLS. The problem is that, due to the simultaneity of τ_t

and π_t, one of the regressors, $\Delta\tau_t$, is correlated with the disturbance term, ε_t. (Foreign inflation π^*_t, is assumed exogenous, and the lagged variables are orthogonal to ε_t as well.) Therefore, an instrumental variable for $\Delta\tau_t \equiv \tau_t - \tau_{t-1}$ is needed. Since τ_{t-1} is already included in the regression, an instrument that is correlated with τ_t but **not** with ε_t will also be adequate. The obvious candidate is g_t, the national income share of gross government expenditure, which is assumed to be weakly exogenous, as in the derivation of the model in Sections 2.2 and 2.3. The whole set of instruments that we use to estimate (2.21) by IV is (1, g, π^*, τ_{-1}, π^*_{-1}, π_{-1}). There is an outlier in the g_t series of Portugal, which is corrected by using a dummy for the year 1980.

In the specification of (2.21) the long-run parameters δ_1 and δ_2 cannot be estimated directly by linear IV; only γ_3 and the products $\gamma_3\delta_1$ and $\gamma_3\delta_2$ can be estimated directly. To obtain estimates of δ_1 and δ_2 one could divide the estimates of $\gamma_3\delta_1$ and $\gamma_3\delta_2$ by the estimated γ_3 (Indirect Least Squares) but the computation of standard errors for these ILS estimates is cumbersome. A more attractive estimator is obtained by rewriting equation (2.21). Replacing π_{t-1} by $\pi_t - \Delta\pi_t$, τ_{t-1} by $\tau_t - \Delta\tau_t$ and π^*_{t-1} by $\pi^*_t - \Delta\pi^*_t$ one obtains the following equation:

$$(2.22) \qquad \pi_t = -\gamma_0/\gamma_3 + \delta_1\tau_t + \delta_2\pi^*_t -$$

$$(\gamma_1/\gamma_3 + \delta_1)\Delta\tau_t - (\gamma_2/\gamma_3 + \delta_2)\Delta\pi^*_t + (\gamma_3 + 1/\gamma_3)\Delta\pi_t + (-1/\gamma_3)\varepsilon_t,$$

which gives direct estimates of δ_1 and δ_2. In equation (2.22) the error term is correlated with **two** right-hand side variables, $\Delta\pi_t$ and τ_t, and therefore OLS yields inconsistent estimates. Note that, from an econometric point of view, both (2.21) and (2.22) correspond to the Euler equation of the tax and seigniorage smoothing model. In the spirit of Hansen's (1982) Generalised Method of Moments, we exploit the orthogonality of ε_t to a set of instruments to estimate the parameters. Because our Euler equations are linear, the GMM estimator of equation (2.22) is equal to the familiar Instrumental Variables estimator, with the same set of instruments used before to estimate (2.21). This IV estimator is numerically identical to the ILS estimator; its advantage is that standard errors are easily computed using the standard expression for the variance-covariance matrix of the IV estimator.

Before presenting the estimates it must be noted that we treat velocity (m^{-1}) as a constant and consequently we do not include it explicitly in the model. The foreign country is Germany, except for Ireland, where it is the UK, and for Germany, where it is the US.

Table 2.2 Estimates of tax/seigniorage smoothing cum EMS-anchor model in terms of inflation rates.

model: $\Delta\pi_t = \gamma_0 + \gamma_1\Delta\tau_t + \gamma_2\Delta\pi^*_t + \gamma_3(\pi_{t-1} - \delta_1\tau_{t-1} - \delta_2\pi^*_{t-1}) + \varepsilon_t$

	γ_1	γ_2	γ_3	δ_1	δ_2	R^2	s.e.	LM
US	-2.055	0.761	-0.172	-0.863	-1.006	-0.59	2.846	2.56
	(0.64)	(1.37)	(0.52)	(0.17)	(0.19)			
UK	1.043	1.216	-0.338	0.558	1.560	0.41	2.990	0.97
	(1.75)	(2.12)*	(1.94)	(1.04)	(1.39)			
BEL	-0.038	0.530	-0.545	0.016	1.789	0.66	1.159	0.06
	(0.06)	(1.95)*	(2.89)*	(0.13)	(3.90)*			
DEN	0.420	0.582	-0.572	0.012	1.522	0.53	1.747	1.40
	(0.67)	(1.58)	(2.62)*	(0.18)	(2.58)*			
FRA	0.701	0.813	-0.447	0.173	1.482	0.49	1.479	0.62
	(1.11)	(2.75)*	(2.08)*	(0.77)	(3.30)*			
GER	-0.418		-0.035	-2.759		-0.04	1.230	2.88
	(0.42)		(0.16)	(0.15)				
ITA	-2.13	-0.04	-0.157	-0.672	5.465	0.34	2.635	0.31
	(1.99)*	(0.05)	(0.82)	(0.25)	(1.28)			
NETH	-0.157	0.737	-0.568	-0.163	1.477	0.50	1.401	1.46
	(0.28)	(2.70)*	(3.08)*	(1.59)	(3.65)*			
GRE	2.666	2.662	-0.778	1.832	2.866	0.28	4.086	0.45
	(0.87)	(1.62)	(2.32)*	(6.12)*	(4.30)*			
IRE	0.917	0.691	-0.625	0.072	0.952	0.66	2.043	0.07
	(1.83)	(5.45)*	(2.81)*	(0.46)	(6.19)*			
POR	-0.217	0.160	-0.521	0.704	3.283	0.44	3.856	2.28
	(0.13)	(0.21)	(2.86)*	(2.99)*	(2.92)*			
SPA	-0.850	0.266	-0.285	-0.105	2.946	0.31	2.786	2.31
	(0.84)	(0.50)	(1.75)	(0.24)	(1.92)			

t-statistics in brackets; *indicates significance at 5% level;
sample period: 1962-1987;
s.e.: standard error of regression; LM: $\chi^2(1)$ Lagrange Multiplier test for first-order serial correlation.

The results presented in Table 2.2 do not generally support the theory. The short-run effect of taxes on inflation (γ_1) is not always positive, and almost nowhere significant. This seems to support the view of Trehan and Walsh (1990) that in the short run other effects dominate the correlation between tax and inflation rates. The estimated value of the catch-up term (γ_3) is always negative, although it is not always significant. This suggests that the error-correction mechanism may not be such a bad description. The long-run effect (δ_1) is significant for only two countries, Greece and Portugal. Precisely these countries relied heavily on seigniorage to finance their budgets in the seventies and the eighties. For the other countries, however, other aspects than seigniorage revenue seem to determine the rate of inflation.

One of those other influences is specified in the model; the desire for exchange rate stability is modelled through the objective of minimising inflation differential with abroad, by the inclusion of the foreign rate of inflation as an explanatory variable. This effect is significant, both in the short run (γ_2) and in the long run (δ_2), for most European countries, except Italy, but not for the US. Germany is excluded since it is considered to be the "foreign" country for the other members of the European Monetary System. This demonstrates that exchange rate policy has been relatively unimportant for the US, whereas it is an important element in macroeconomic policy for most European countries, except perhaps Italy. The reason that exchange rate stability vis-à-vis the Deutschemark did not show up significantly for Italy may be due to the presence of capital controls.

Table 2.3 Estimates of tax/seigniorage smoothing cum EMS-anchor model in terms of nominal interest rates.

model: $\Delta i_t = \gamma_0 + \gamma_1 \Delta\tau_t + \gamma_2 \Delta i^*_t + \gamma_3 (i_{t-1} - \delta_1 \tau_{t-1} - \delta_2 i^*_{t-1}) + \varepsilon_t$

	γ_1	γ_2	γ_3	δ_1	δ_2	R^2	s.e.	LM
US	-0.302	0.755	-0.223	0.173	0.507	0.55	1.669	0.01
	(0.26)	(4.22)*	(0.91)	(0.10)	(0.52)			
UK	1.844	0.059	-0.814	0.862	-0.407	0.36	2.327	2.26
	(1.95)	(0.20)	(2.11)*	(3.63)*	(0.91)			
BEL	1.771	0.518	-1.148	0.384	0.426	0.34	1.944	3.84
	(0.91)	(2.64)*	(2.59)*	(3.74)*	(1.70)			
DEN	-0.269	0.363	-0.362	0.464	0.490	0.23	2.730	3.52
	(0.38)	(1.22)	(2.13)*	(2.67)*	(0.61)			
FRA	0.981	0.656	-0.609	0.354	1.057	0.77	1.118	0.17
	(2.72)*	(5.95)*	(3.52)*	(3.37)*	(5.10)*			
GER	-0.952	0.844	-0.459	-0.524	1.037	0.61	1.597	0.03
	(0.98)	(4.23)*	(1.62)	(1.25)	(2.45)*			
ITA	0.905	0.688	-0.337	0.249	1.632	0.31	1.928	0.04
	(1.00)	(2.20)*	(1.65)	(0.47)	(2.32)*			
NETH	0.198	0.703	-0.680	0.126	0.398	0.68	1.461	5.61*
	(0.20)	(4.71)*	(3.11)*	(1.15)	(1.43)			
GRE	0.776	0.298	-0.400	1.025	0.783	0.21	1.089	2.48
	(1.07)	(1.50)	(1.71)	(5.72)*	(2.20)*			
IRE	-0.036	1.229	-0.874	0.348	1.299	0.47	2.834	2.42
	(0.04)	(3.37)*	(3.32)*	(1.43)	(2.93)*			

Notes see Table 2.2; sample period for Greece is 1964-1986.
For Portugal and Spain no good interest rate data were available.

A variant on model (2.21) or (2.22) is a relation between nominal interest rates and tax rates. Table 2.3 shows the results of that model. The results are a bit confusing; the "tax" effect becomes more significant for the UK, Belgium, Denmark and France, and remains important for Greece. The effects of foreign nominal interest rates seems not to be as clear as the effects of foreign inflation. The effects seem to contradict the results of the EMS-anchor model, especially for Italy and the Netherlands. This can be explained in two ways: either capital controls have been important (Italy) or the central bank of the country under concern has a better reputation and discipline than the Bundesbank (the Netherlands).

Table 2.4 Estimates of tax/seigniorage tradeoff with unemployment.

model: $\Delta\pi_t = \gamma_0 + \gamma_1\Delta\tau_t + \gamma_2\Delta\pi^*_t + \gamma_3\Delta u_t + \gamma_4(\pi_{t-1} - \delta_1\tau_{t-1} - \delta_2\pi^*_{t-1} - \delta_3 u_{t-1}) + \varepsilon_t$

	γ_1	γ_2	γ_3	γ_4	δ_1	δ_2	δ_3	R^2	LM
US	-0.844	0.355	-2.439	0.005	-228.4	-91.3	321.1	0.53	0.15
	(0.58)	(0.98)	(1.98)*	(0.02)	(0.2)	(0.2)	(0.2)		
UK	1.559	0.669	-1.761	-0.264	2.412	1.748	-2.100	0.52	0.31
	(2.88)*	(1.09)	(2.28)*	(1.36)	(1.42)	(1.02)	(1.22)		
BEL	0.378	0.423	-0.37	-0.603	0.137	1.740	-0.115	0.74	0.08
	(0.20)	(0.66)	(0.42)	(1.47)	(0.21)	(1.93)	(0.16)		
DEN	0.129	0.649	0.61	-0.745	-0.011	0.948	0.169	0.56	1.56
	(0.25)	(1.61)	(1.63)	(2.99)*	(0.14)	(2.27)*	(0.67)		
FRA	0.539	0.806	-0.11	-0.370	-0.491	1.620	0.827	0.50	0.07
	(0.77)	(2.59)*	(0.10)	(1.48)	(0.28)	(2.53)*	(0.41)		
GER	-0.967	-0.032	-0.601	0.025	9.268	-12.02	0.866	0.12	1.99
	(1.00)	(0.18)	(1.14)	(0.06)	(0.07)	(0.07)	(0.03)		
ITA	-1.908	0.024	-1.820	-0.146	-1.312	6.323	3.820	0.51	0.56
	(1.53)	(0.03)	(1.21)	(0.74)	(0.19)	(1.09)	(0.35)		
NETH	0.416	0.362	-0.371	-0.810	0.191	1.069	-0.542	0.63	0.95
	(1.10)	(1.22)	(0.96)	(3.99)*	(1.42)	(3.68)*	(2.28)*		
GRE	3.609	3.026	0.355	-0.909	2.093	2.488	-0.507	0.17	0.00
	(0.93)	(1.48)	(0.26)	(2.11)*	(4.75)*	(2.53)*	(0.66)		
IRE	1.355	0.600	-0.059	-0.916	0.701	0.702	-0.696	0.67	0.21
	(2.38)*	(3.99)*	(0.10)	(3.31)*	(2.42)*	(5.27)*	(2.14)*		
POR	1.929	0.392	-2.479	-0.281	1.638	4.746	-3.113	0.25	0.60
	(0.60)	(0.30)	(0.62)	(0.50)	(0.33)	(0.67)	(0.21)		
SPA	0.816	-0.119	0.023	-0.523	1.914	1.055	-1.509	0.41	1.07
	(0.73)	(0.18)	(0.03)	(2.42)*	(1.89)	(1.10)	(1.91)		

Notes see Table 2.2.

The results of estimating the model with unemployment, based on equation (2.19), are presented in Table 2.4. In general, unemployment does not seem to be an important determinant of inflation in the way predicted by the model of Section

2.5. In fact, for the US and the UK changes in unemployment exert a significant downward pressure on inflation rates, whereas the model predicts a positive effect. In the long run unemployment does not seem to have a big impact on the rate of inflation, except in the Netherlands, Ireland and Spain, where there is a negative long-run effect of the unemployment rate. The effects of taxes and foreign inflation do not seem to be affected much by the inclusion of unemployment in the tax-seigniorage equation.

2.7 Non-nested tests against cost-push theories of inflation

One wonders how well the model of tax and seigniorage smoothing compares with alternative models of inflation in the EEC, especially in the seventies. The tax/seigniorage smoothing model treats inflation, money supply or nominal interest rates, as policy instruments, i.e. variables that the government can manipulate at will. However, the demand for money function is notoriously unstable and a variety of external and internal shocks hit the economy continually. One may well have doubts that under such circumstances inflation is the result of deliberate government policy to smooth seigniorage revenues and obtain an optimal tax-inflation tradeoff.

According to Bruno and Sachs (1985) the high inflation spells of the 1970s must be primarily explained by raw materials price shocks and the subsequent upward wage-price spiral, caused by pressure to keep real wages at an unrealistically high level. The spiral is only brought to rest when monetary authorities take severe measures to stop inflation, thereby forcing a slowdown in activity in the real economy and causing high unemployment. Such a view of inflation is better captured by an augmented Phillips curve. It differs from the traditional Phillips curve in that it takes expected inflation into account, but also unexpected external price shocks and various supply side variables, such as productivity growth and the level of real wages relative to the full-employment level (the so-called "wage gap"), play a role. Bruno and Sachs do not account explicitly for the changes in prices of imported goods other than of raw materials. For the smaller European countries, however, imported inflation may be very significant, because their economies are very open and a large proportion of consumption goods is imported. This affects inflation, since the measure used is the relative change in the consumer price index. This is also the reason why another important source of imported inflation is depreciation of the exchange rate. We do not take this as a separate effect, but include it in the raw materials prices and the prices of imports. Hence, raw material price inflation (π^R) is the relative price change of raw materials in US dollars (π^{WP}) plus the rate of depreciation of the currency against the dollar ($\dot{e}_{US}$). Similarly, inflation in import prices (π^M) is defined as the foreign inflation rate (π^*) plus the rate of depreciation of the nominal exchange rate ($\dot{e}$). The estimation results of such a cost-push model of inflation are presented in Table 2.5.

Table 2.5 Estimates of cost-push model of inflation.

$$\pi_t = \beta_0 + \beta_1 D_t + \beta_2 \pi_{t-1} + \beta_3 \pi^R_t + \beta_4 \pi^R_{t-1} + \beta_5 \pi^M_t + \beta_6 \pi^M_{t-1} + \beta_7 u_t + \beta_8 \pi_{t-2} + \varepsilon_t$$

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	R^2
US	2.394	0.842	0.058	0.062		-0.867			0.95
	(4.33)	(12.24)	(4.27)	(4.27)		(4.88)			
UK	7.524	0.445		0.071	0.089	-0.101	-0.940		0.80
	(2.89)	(2.90)		(1.52)	(1.00)	(0.96)	(3.22)		
BEL	3.240	0.473		0.050	0.230		-0.390		0.88
	(3.03)	(4.22)		(2.82)	(3.68)		(3.47)		
DEN	3.366	0.314	0.052	0.047	0.283		-0.418		0.78
	(2.51)	(2.04)	(1.96)	(1.86)	(2.65)		(2.16)		
FRA	4.013	0.908	0.073	0.022			-0.190	-0.570	0.90
	(3.48)	(5.37)	(3.84)	(1.16)			(1.24)	(2.89)	
GER	0.817	0.736	0.034				-0.312		0.77
	(1.12)	(5.75)	(2.10)				(2.23)		
ITA	6.260	0.393	0.087	0.029	0.024	0.031	-0.820		0.91
	(2.27)	(2.57)	(2.17)	(0.69)	(0.20)	(0.22)	(1.88)		
NETH	0.650	0.557	0.020	0.010	-0.040	0.334	-0.258		0.76
	(0.52)	(3.09)	(0.67)	(0.35)	(0.20)	(1.79)	(1.73)		
GRE	5.040	0.295	0.140	0.115	0.193		-0.122		0.90
	(1.59)	(1.42)	(2.73)	(2.14)	(1.76)		(0.47)		
IRE	1.127	0.228	0.042	0.026	0.324	0.170	-0.003		0.89
	(0.53)	(1.07)	(1.30)	(0.80)	(3.20)	(1.25)	(0.01)		
POR	10.041	0.206	0.138	0.143	0.101	-0.121	-0.654		0.94
	(4.80)	(1.29)	(3.43)	(3.39)	(1.15)	(1.33)	(1.54)		
SPA	5.296	0.415	0.013	0.081	0.123	-0.071	-0.275		0.83
	(2.26)	(2.26)	(0.31)	(1.95)	(1.33)	(0.75)	(2.44)		

Notes: see Table 2.2; method of estimation: OLS;
sample period for Greece, Portugal and Spain: 1964-1987.
D7487: dummy variable for 1974-1987 period.

In this model, the slowdown in productivity growth and the increase in the wage gap resulting from the first oil shock is proxied by a dummy variable that is zero before 1974 and one from 1974 onwards. Following Bruno and Sachs (1985), expected inflation is proxied by lagged inflation. Because the effect of raw materials price shocks and imported inflation sometimes appeared to occur with a lag, both current and lagged π^R and π^M are included in the regression equation. The equations are usually free of serial correlation, but for France π_{t-2} is included as a regressor to get rid of serial correlation. Some very insignificant regressors were omitted from the equations.

The results indicate that the cost-push model explains the inflation for most countries quite well. The R^2 of most equations is fairly high, and for each country some significant effects show up. The effect of lagged inflation is significant, but the coefficient is smaller than one for all countries. This means that there is some delay in the adjustment of prices to the external shocks, perhaps due to staggered wage contracts as in Taylor (1979) or Alogoskoufis (1990). There is a clear trend-break in 1974 for all countries except Germany, the Netherlands and Ireland.

The timing of the effect of the raw materials price shocks is not the same for all countries, sometimes π^R_t is significant, sometimes π^R_{t-1} and sometimes both, but it is clear that raw material price changes are important determinants of inflation in all countries. In this respect also the sharp rise and subsequent fall of the dollar around 1985 is important for the European countries since most raw materials prices are expressed in US dollars. Imported inflation seems to have been important especially for the smaller European countries, Belgium, Denmark, the Netherlands, Greece and Ireland, and to a lesser extent Portugal and Spain. The larger countries, the UK, France and Italy and the US appear to be less sensitive to import price fluctuations. For all countries there is a significant Phillips curve effect. High unemployment exerts a downward pressure on inflation, only for Ireland and Greece the effect is rather weak.

We specified and estimated the cost-push model of inflation primarily as a simple alternative to the tax/seigniorage smoothing cum EMS-anchor model of inflation. The empirical specifications of models (2.21) and (2.23) are not nested, so we cannot perform a direct test of one model against the other by examining the significance of one or more explanatory variables. We perform the test of non-nested hypotheses proposed by Godfrey and Pesaran (1983) on the cost-push model (M1) against the tax/seigniorage smoothing cum EMS-anchor model (M2). To carry out the test, M2 is specified as a simple regression of π_t on a constant, π_{t-1}, τ_t, τ_{t-1}, π^*_t and π^*_{t-1}. This is not fully correct, as τ_t is endogenous and should be instrumented. However, we do not know of any test of non-nested simultaneous equations. Therefore, for purposes of carrying out the non-nested tests, we estimate both models here by OLS.

Table 2.6 Non-nested tests of cost-push model (M1) versus tax/seigniorage cum-EMS anchor model (M2).

country	M1 vs. M2	M2 vs. M1
US	-1.56	-11.92*
UK	-1.72	-3.52*
Belgium	-3.38*	-2.09*
Denmark	-0.43	-2.86*
France	-2.37*	-3.49*
Germany	-0.73	-0.85
Italy	-3.38*	-3.39*
Netherlands	-2.17*	-0.45
Greece	0.03	-3.21*
Ireland	-2.86*	-1.63
Portugal	-1.30	-7.13*
Spain	-0.04	-3.59*

* denotes rejection at 5% level.

The results in Table 2.6 show that in most cases the tax/seigniorage smoothing cum EMS-anchor model is rejected against the cost-push model. Exceptions to this rule are Germany, the Netherlands and Ireland. For the latter two countries the cost-push model of inflation is even rejected against the tax/seigniorage smoothing cum EMS-anchor model. Especially the objective of maintaining a stable exchange rate with Germany and the UK, respectively, has been the dominant determinant of inflation for these two countries. This is not surprising, since the Netherlands has been one of the most loyal members of the European Monetary System and until Ireland joined the European Monetary System it had been tracking the pound sterling very closely. For Germany the outcome is undecided, the data cannot make a clear choice between the two models. The cost-push model clearly is a better model of inflation for the US, the UK, Denmark, Greece, Portugal and Spain. The conclusion for Portugal and Greece is somewhat unexpected, because Table 2.2 suggests that the tax effect was important for these countries. For Belgium, France and Italy the test rejects both models against each other, indicating that we have to look for a better model of inflation for those countries.

2.8 Concluding remarks

The tax and seigniorage smoothing model has been extended to allow for an exchange rate anchor and stabilisation policy. An analysis of the time inconsistency of monetary policy arising from the use of a surprise inflation tax to wipe out the real value of debt illustrated the gains from membership of the European Monetary System. By tying one's hands to the policy of the Bundesbank, one obtains the discipline of the Bundesbank. These advantages may outweigh the disadvantages in terms of a sub-optimal government revenue mix when there is a large outstanding stock of nominal government debt and when the system of tax collection is quite efficient. The exchange rate anchor seems to explain a lot of the inflation in EMS countries. We find little evidence in favour of co-integration of taxation and seigniorage revenues. A comparison of the revenue-smoothing model with a cost-push model of inflation clearly prefers the latter model for most countries, except those that have pegged their currencies firmly to a foreign currency (the Netherlands and Ireland).

Data appendix.

General government expenditure as a fraction of national income (including interest payments), g. Source: OECD Economic Outlook, various issues.

General government revenues as a fraction of national income, τ. Source: OECD Economic Outlook, various issues.

Standardised unemployment rate, u. Source: OECD Economic Outlook, various issues.

Consumer price index, P. Source: International Financial Statistics Yearbook 1990 (IFS) line 64.

Short term interest rate, i. Source: IFS line 60b, except UK: line 60c; Italy, Ireland, Spain: OECD Main Economic Indicators.

Money base, M. Source: IFS line 14.

National Income, Y. Source: IFS line 99b (GDP), except US, Belgium, Germany and the Netherlands, IFS line 99a (GNP).

Exchange rate of US dollar, e. Source: IFS line ahx.

Commodity price index in US dollars, R. Source: IFS line d (pp. 182/183).

Debt, D. Source: IFS lines 88 and 89, various national statistics[4].

Inflation, $\pi \equiv \Delta\ln(P)$.

Government liabilities, $m + d = (M+D)/Y$.

Depreciation of exchange rate, $\dot{e} = \Delta\ln(e)$.

Raw materials price shock (in US dollars), $\pi^{R}_{US} = \Delta\ln(R)$.

Raw materials price shock (in domestic currency), $\pi^{R} = \pi^{R}_{US} + \dot{e}$.

Imported inflation, $\pi^{M} = \pi^{*} + \dot{e}$.

4 The government debt data for Ireland were kindly provided by Jeroen Kremers of the Dutch Ministry of Finance.

Chapter 3

A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model

3.1 Introduction.

Following the seminal model of Krugman (1991), a large theoretical literature of exchange rate determination in target zones has developed. The crucial observation in these models is that the target zone influences the expectations of future spot prices because the central banks will intervene if the exchange rate deviates too much from its central parity. Hence, in a forward looking model of exchange rate determination, the presence of a target zone has an impact on the exchange rate itself, even if there are currently no interventions. One of the most striking implications of the Krugman model, and nearly all other theoretical models, is the non-linear, S-shaped relation between fundamentals and exchange rates.

In this paper, we develop several methods for estimating the Krugman (1991) target zone model. For this model the likelihood function is known, though quite complicated, so efficient estimation and testing based on the method of Maximum Likelihood is possible. However, for almost any extension of the model that is present in the literature, the likelihood function is not known, so other estimators are needed. An alternative class of estimators is given by the Generalised Method of Moments. A recent development in GMM estimation is the Method of Simulated Moments, where the model is simulated to compute the moments of the theoretical distribution. Although the computational burden of simulating artificial data and computing their moments is high, simulated moments estimators seem well suited for applications in target zone models, because in these models the stochastic process generating fundamentals and exchange rates is explicitly specified, and simulation of the model is conceptually straightforward.

In the empirical part of this chapter the Krugman target zone model is estimated and tested on EMS exchange rate data over a relatively stable period, January 1987 to October 1990. The results indicate that there are significant non-linearities in the exchange rate processes of the Belgian franc, the French franc and the Danish krone, possibly caused by the impact of the band on expectations of future variables. For the other currencies, however, there seem to be no non-linearities at all, which result may be explained by the presence of an implicit band that is narrower than the official EMS band. The specification of the model is tested by comparing some moments of the theoretical distribution with the sample moments. These tests reveal that the Krugman target zone model is misspecified: the model is not capable of explaining the full magnitude of the observed conditional heteroskedasticity and leptokurtosis in exchange rate returns.

The organisation of the chapter is as follows. Section 2 reviews the Krugman target zone model briefly. Section 3 discusses Maximum Likelihood estimation, and Section 4 Method of Simulated Moments estimation of the Krugman target zone model. Section 5 presents the empirical results and Section 6 concludes the chapter. Two Appendices give details on the simulation of the model and the computation of the specification tests.

3.2 A simple model of exchange rate determination in a target zone.

In this section we describe the basic target zone model of Krugman (1991). In this model, the logarithm of the exchange rate, e(t), is a function of the economic fundamental and its own expected rate of change

$$(3.1) \qquad e(t) = f(t) + \alpha \cdot E_t(de(t)/dt), \qquad \alpha > 0.$$

Excluding 'bubble' solutions, this implies that e(t) is the present discounted value of all expected future fundamental values

$$(3.2) \qquad e(t) = \alpha^{-1} \int_0^\infty \exp(-s/\alpha) E_t[f(t+s)] ds.$$

It is assumed that, except for occasional interventions, the fundamental follows a Brownian motion with constant drift, μ, and variance, σ^2,

$$(3.3) \qquad df(t) = \mu dt + \sigma dW(t),$$

where W(t) is the Wiener process with standard normal increments over the unit time interval. The Krugman model assumes a very specific intervention policy: f(t) follows (3.3) until a lower limit $\underline{f}$ or an upper limit $\bar{f}$ is reached. If so, the fundamental is kept within the band by infinitesimal interventions at the margin. This stochastic process is called a regulated Brownian motion with reflecting barriers, see Harrison (1985).

The regulated Brownian motion exhibits a *strong Markov property* (Harrison, 1985, p.81) so that the expectations of $f(t+s)$, $s>0$, depend only on the current value of $f(t)$. Therefore, $e(t)$ can be written as a function of $f(t)$ only, say $e(t)=G(f(t))$. Using Ito's lemma, $G(f(t))$ is know to satisfy the following differential equation:

$$(3.4) \qquad dG(f(t)) = [\mu G'(f(t)) + \tfrac{1}{2}\sigma^2{}''(f(t))]dt + G'(f(t))\cdot\sigma dW(t).$$

Of course, the exchange rate solution also has to satisfy (3.1). Substituting (3.4) into (3.1) one obtains

$$(3.5) \qquad G(f(t)) = f(t) + \alpha[\mu G'(f(t)) + \tfrac{1}{2}\sigma^2 G''(f(t))].$$

This differential equation should be satisfied by the function $G(f(t))$. The general solution for $G(f(t))$ is

$$(3.6) \qquad G(f(t)) = f(t) + \alpha\mu + A_1\exp(\lambda_1 f(t)) + A_2\exp(\lambda_2 f(t)),$$

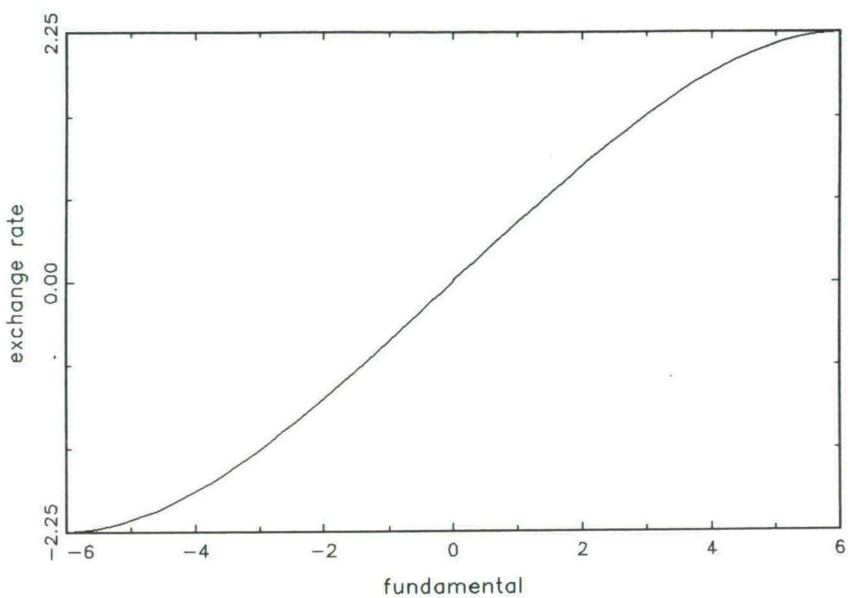
where λ_1 and λ_2 are the roots of the characteristic equation

$$(3.7) \qquad \tfrac{1}{2}\alpha\sigma^2\lambda^2 + \alpha\mu\cdot\lambda = 1,$$

and A_1 and A_2 are constants that are to be determined from the relevant boundary conditions provided by the target zone model. If the exchange rate regime were a free float, the exclusion of 'bubble' solution implies that $A_1=A_2=0$ and hence the solution would be linear, $e(t)=f(t)+\alpha\mu$.

In the model of Krugman (1991) there is a fully credible target zone $[\underline{e},\bar{e}]$ with infinitesimal interventions at the margin. The boundary conditions for the Krugman target zone model are the **value matching** and the so-called **smooth pasting** conditions. Svensson (1992) and Bertola (1993) give an intuitive motivation of these conditions. With infinitesimal marginal interventions the value matching conditions are $G(\bar{f})=\bar{e}$ and $G(\underline{f})=\underline{e}$. The smooth pasting conditions are that the derivative of the exchange rate function $G(f(t))$ at $\underline{f}$ and $\bar{f}$ should be zero: $G'(\underline{f})=0$ and $G'(\bar{f})=0$. Using these conditions, the solution for the function $G(f(t))$ is non-linear and has an S-shaped form. Note that in solving the model the value matching and smooth pasting conditions have to be solved simultaneously in the four unknown parameters $\underline{f}$, $\bar{f}$, A_1 and A_2. Figure 3.1 graphs the exchange rate solution.

Figure 3.1: Solution of Krugman target zone model



3.3 Maximum Likelihood estimation.

In this section we show how the Krugman (1991) target zone model can be estimated efficiently by the method of Maximum Likelihood (ML). In order to apply ML, the statistical distribution of a sequence $\{e_1(\theta),..,e_T(\theta)\}$ generated by the model must be known to derive the likelihood function $L(\theta)$, defined as the joint density[1] of the observations $D(e_1,..,e_T|\mathbf{E}_0;\theta)$, where $\mathbf{E}_0$ is the set of initial conditions. Using the prediction error decomposition, the joint density is written as the product of conditional densities

[1]The symbol D is used to denote any (joint) density function.

$$(3.8) \qquad L(\theta) \equiv D(e_1,..,e_T|\mathbf{E}_0;\theta) = \prod_{t=1}^{T} D(e_t|\mathbf{E}_{t-1};\theta),$$

where $\mathbf{E}_{t-1}$ denotes the information set formed by the initial conditions $\mathbf{E}_0$ and all observations up to and including e_{t-1}.

In the Krugman target zone model the fundamental follows a regulated Brownian motion on the interval $[\underline{f},\bar{f}]$. It is therefore convenient to rewrite the likelihood function in terms of the fundamentals. The mapping $e=G(f;\theta)$ is monotone and continuously differentiable for any θ so we can apply a change of variables, $f_t=G^{-1}(e_t;\theta)$, for all observed values of the exchange rate and rewrite the likelihood function to

$$(3.9) \qquad L(\theta) = \prod_{t=1}^{T} D(f_t|\mathbf{E}_{t-1};\theta)\cdot G'(f_t;\theta)^{-1}$$

The latter part of (3.9) is the Jacobian of the transformation, which is always positive and continuous. Due to the strong Markov property of the regulated Brownian motion, the history is irrelevant for the distribution of the future of the process if f(t) is known. This property allows us to condition the distribution of f_t on the previous observation only, and the likelihood function can be rewritten to

$$(3.10) \qquad L(\theta) = \prod_{t=1}^{T} D(f_t|f_{t-1};\theta)\cdot G'(f_t;\theta)^{-1}$$

In our empirical work we condition on the the first observation, e_0, which is transformed to the initial state of the fundamental process through $f_0=G^{-1}(e_0;\theta)$.

Having derived the likelihood of the exchange rate in terms of the fundamentals, what remains to be found is the predictive density function of the regulated Brownian motion. The predictive distribution function of the *one-sided* regulated Brownian motion with regulation at lower bound $\underline{f}$ is given in Harrison (1985, p.49). The expression is

$$(3.11) \qquad P(f|f_{t-s}) = \Phi\left(\frac{f-f_{t-s}-\mu s}{\sigma\sqrt{s}}\right) - e^{\tau(\underline{f}-f)}\Phi\left(\frac{2\underline{f}-f-f_{t-s}-\mu s}{\sigma\sqrt{s}}\right), \quad \tau\equiv 2\mu/\sigma^2$$

The first part of this function is the usual normal distribution function, whereas the second part represents the probability that f(t) is regulated at the lower bound in the time interval (t-s,t]. The marginal distribution is obtained by letting s go to infinity which yields

$$(3.12) \qquad P(f) = \begin{cases} 1-e^{\tau(f-\underline{f})} & \text{if } \tau<0 \\ 0 & \text{if } \tau \geq 0 \end{cases}$$

If the target zone were one-sided, one could apply these distribution functions directly to obtain the likelihood function. However, most actual target zones are two-sided, so we need the conditional density or distribution function of a *two-sided* regulated Brownian motion. This function is quite complicated[2] and contains an infinite summation, which makes computation very time-consuming.

Instead of using this exact density we approximate the conditional distribution of a two-sided regulated Brownian motion by a weighted average of the conditional distributions of two one-sided regulated Brownian motions, regulated at the lower and at the upper bound, respectively. The weights are chosen such as to satisfy two conditions. First, the approximate conditional distribution must converge to the exact marginal distribution as the time between f(t) and the initial value f(0) goes to infinity. Second, the function must converge to the predictive distribution of a one-sided regulated Brownian motion if one of the bounds goes to infinity. The distribution function that satisfies these conditions is

---

[2] The density of a regulated Brownian motion f(t) with drift μ, variance σ^2 and support $[\underline{f},\bar{f}]$, conditional on $f(0) = f_0$ is

$$p(f|f_0) = \frac{\tau e^{\tau(f-\underline{f})}}{e^{\tau(\bar{f}-\underline{f})}-1} + \frac{\exp[\tau(f-f_0)/2]}{4(\bar{f}-\underline{f})} \sum_{n=1}^{\infty} \frac{y_n(f)\cdot y_n(f_0)}{\lambda_n a^2/\sigma^2} \cdot \exp(-\lambda_n t)$$

$\tau \equiv 2\mu/\sigma^2$

$a \equiv (\bar{f}-\underline{f})/\pi$

$y_n(f) \equiv 2n\cdot\cos(n(f-\underline{f})/a) + \tau a\cdot\sin(n(f-\underline{f})/a)$

$\lambda_n \equiv \sigma^2\left[n^2/a^2 + \tau^2/4\right]/2$

These formulas are adapted and corrected from Appendix A3 of the working paper version of Svensson (1991b).

$$P(f|f_{t-s}) = \Phi\left(\frac{f-f_{t-s}-\mu s}{\sigma\sqrt{s}}\right) - (1-P(f))\Phi\left(\frac{2\underline{f}-f-f_{t-s}-\mu s}{\sigma\sqrt{s}}\right) + P(f)(1-\Phi\left(\frac{2\bar{f}-f-f_{t-s}-\mu s}{\sigma\sqrt{s}}\right)), \tag{3.13}$$

where

$$P(f) = \frac{e^{\tau(f-\underline{f})}-1}{e^{\tau(\bar{f}-\underline{f})}-1} \quad \text{if } \tau \neq 0, \qquad P(f) = \frac{f-\underline{f}}{\bar{f}-\underline{f}} \quad \text{if } \tau = 0,$$

is the marginal or 'asymptotic' distribution of f(t), see Harrison (1985, p.90). Numerical experimentation shows that the first derivative of this approximate distribution function gives a very accurate approximation to the exact density function.

In order to test the accuracy and normality of the ML estimator in a reasonably large sample we picked a vector of pseudo-true parameter values and generated a number of artificial time series of exchange rates, using the simulation method of the target zone model described in the Appendix. Each simulated series contains 1000 observations. The pseudo-true parameter values picked are $(\mu,\sigma^2,\alpha)=(0,4,0.1)$ and the exchange rate is allowed to fluctuate in a band with bounds $-\underline{e}=\bar{e}=2.25$; these values imply first and second moments for the exchange rate that are comparable to those of actual EMS exchange rates. For each simulated series the parameter vector was estimated by ML using the approximate likelihood function. In Figures 3.2-3.4, the empirical distribution functions of 100 estimates of μ, $\ln(\sigma^2)$ and $\ln(\alpha)$ are plotted against a normal distribution with mean equal to the pseudo-true parameter value and variance equal to the variance of the estimates. The ML estimator appears to be normal with the correct mean, although there is a slight overestimation of the variance of the fundamental. The variance of the estimates of μ and $\ln(\alpha)$ is rather big, indicating that these parameters are not very precisely estimated.

According to Chernoff (1954), the Likelihood Ratio test of $H_0{:}\alpha=0$ against $H_1{:}\alpha>0$ has a $\frac{1}{2}\chi_0^2+\frac{1}{2}\chi_1^2$ distribution. This distribution function and the empirical distribution function computed from 100 Monte Carlo replications with pseudo-true parameter values $(\mu,\sigma^2,\alpha)=(0,4,0)$ is shown in Figure 3.5. The empirical distribution lies everywhere to the left of the theoretical one, so that using the critical values of Chernoff's distribution gives a conservative test.

Figure 3.2: Empirical distribution ln(alpha)

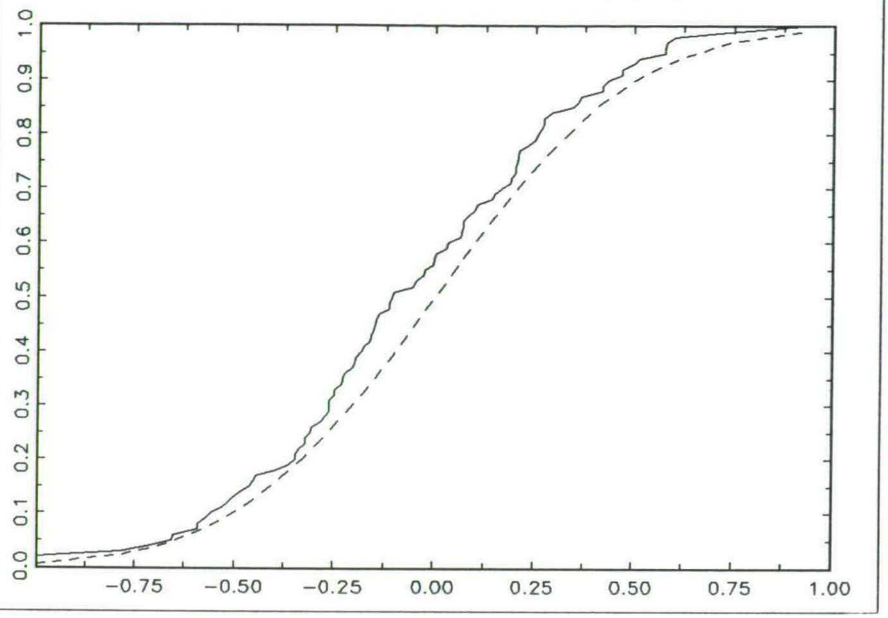


Figure 3.3: Empirical distribution mu

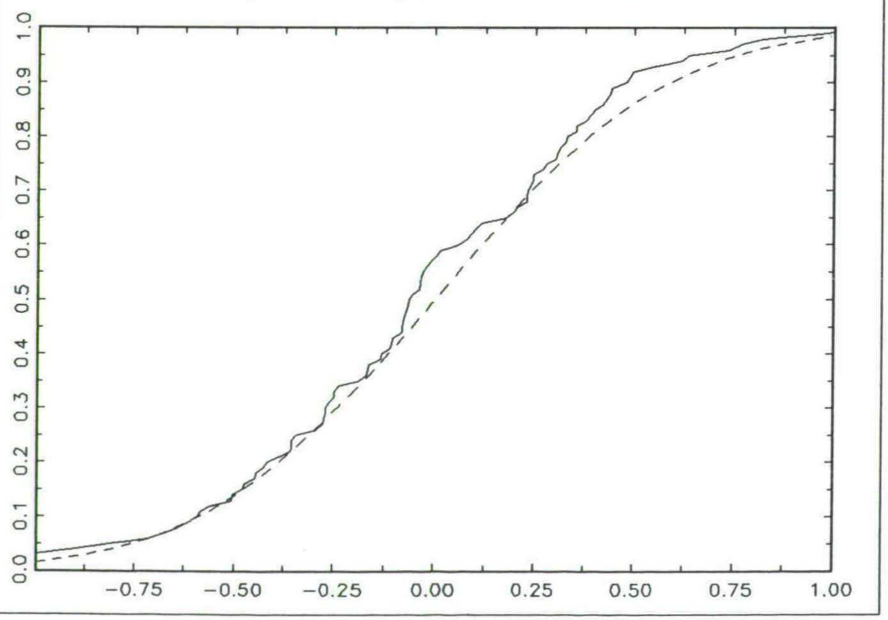


Figure 3.4: Empirical distribution ln(sigma2)

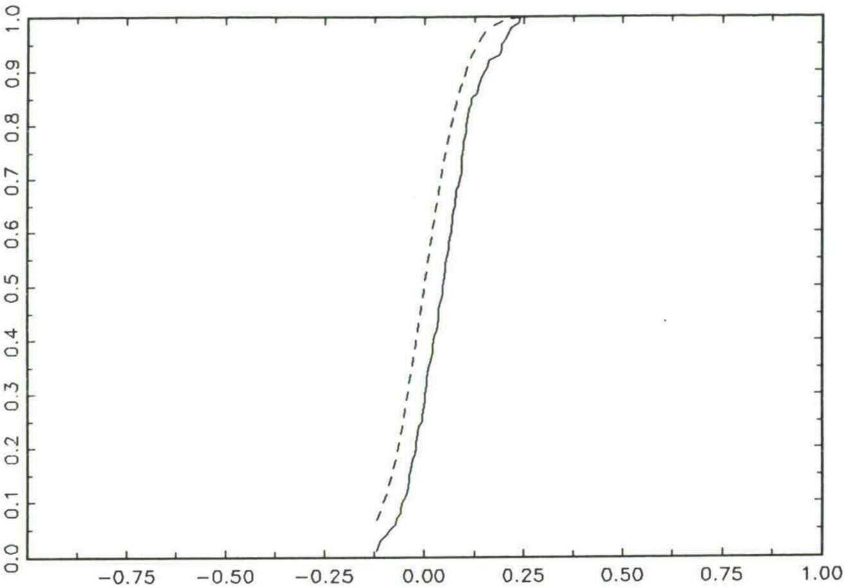
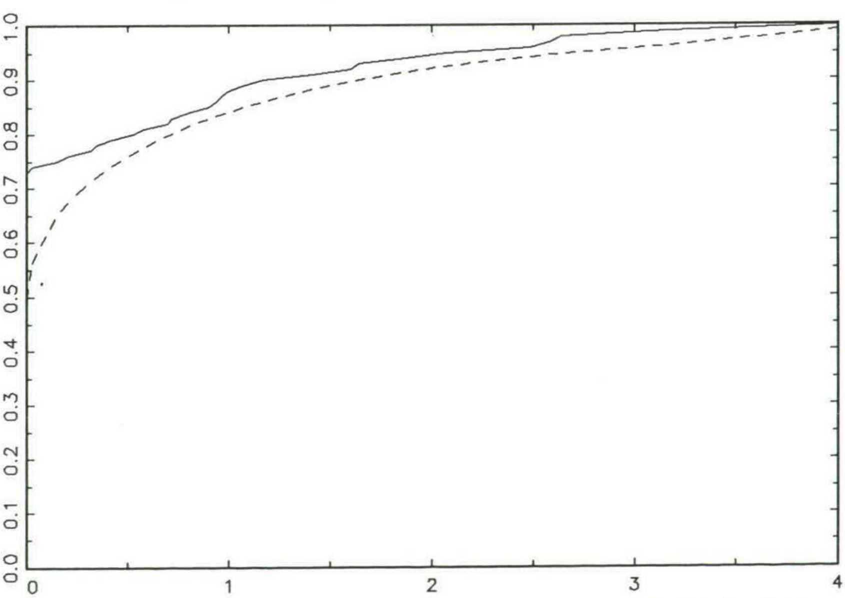


Figure 3.5: Empirical distribution of LR statistic



3.4 Method of Simulated Moments estimation of target zone models.

In the Krugman model the marginal and conditional densities of the exchange rate are known so that Maximum Likelihood estimation of the parameters is feasible. For more complicated target zone models the marginal and conditional exchange rate distributions are usually unknown. An estimator for these models is provided by the method of moments, in which the parameters are chosen such as to minimise the distance between a vector of observed (sample) moments and the theoretical (population) moments. However, in the target zone models computation of some useful moments, especially the autocovariances of the first differences (returns) of the exchange rates, requires multidimensional integration which may be problematic. McFadden (1989) and Lee and Ingram (1991) show that this problem can be solved by computing the theoretical moments by simulation. This method is especially attractive for application to target zone models because the stochastic process that drives the exchange rate is usually explicitly specified in these models.

We compute the moments of the exchange rate and its first difference as follows. A number of series $\{\tilde{e}_{it}\}$ of length equal to the sample size are simulated according to the procedure described in Appendix A. For each simulated series its 'sample' moments are computed, and a consistent estimator of the population moments is obtained by averaging the 'sample' moments over all simulated series. Because our sample starts just after a realignment there is no obvious reason why the first observation would come from the marginal distribution. Therefore, the analysis is performed conditional on the first observation, i.e. the simulation of each series is started at the value of the first observation.

The Method of Simulated Moments (MSM) estimator involves minimising the distance between the sample moments and the simulated population moments, denoted by $d(\theta)$, with respect to the parameters of the model. The most efficient MSM estimator is obtained if one uses the inverse of the covariance matrix of the moments, Σ, to define the metric in which the distance is minimised. We estimate Σ by the method of Newey and West (1987) from the simulated values. Lee and Ingram (1991) show that if the simulated exchange rates are independent of the observed series the asymptotic covariance matrix of the distance vector is $(1+\frac{1}{n})$ times the variance of the moments, where n is the number of simulated exchange rate series. In our applications $n=50$ so that the variance caused by simulating the moments is relatively small. They also show that the asymptotic distribution of the MSM estimator is

$$(3.14) \qquad \sqrt{T}(\hat{\theta}-\theta_0) \underset{asy}{\sim} N\left(0,\ (1+\frac{1}{n})(D'\Sigma^{-1}D)^{-1}\right), \qquad D \equiv \partial d(\theta)/\partial\theta.$$

In actual applications of the method of moments we want to use moments that are most informative about the parameters to be estimated. Although the exchange rate is a stationary process in the target zone model the moments of the *level* of the exchange rate are not very informative about the parameters, the reason being that the observations on the levels exhibit strong serial correlation and therefore very long time series are needed to obtain precise estimates of the theoretical moments of the level of the exchange rate[3]. Moreover, not all parameters of the target zone model can be estimated from the moments of the marginal distribution of the exchange rate alone, because the function G(.), and the marginal distribution of the fundamental depend only on the ratio μ/σ^2 and the product $\alpha\mu$. This argument also implies that histograms of the marginal distribution of the exchange rate (e.g. used by Bertola and Caballero (1992)) alone do not provide information about all the properties of the model. Information from the conditional distribution of the exchange rate is necessary for estimating all parameters. Therefore, it is necessary to use moments of the first differences of the exchange rate. These moments have the additional advantage that their variance is relatively small because first differences of the exchange rate are not strongly correlated over time.

In our empirical application we use five moments of the first difference of the exchange rate. The first two moments are the expectation and the variance, because these are likely to be sensitive to the drift parameter, μ, and the variance parameter, σ^2, of the fundamental process. In addition, we use the third moment (skewness) and the first and second autocovariances of the first difference of the exchange rate. These moments are included to capture the non-linear part of the exchange rate process. In principle a lot of other moments could be added to this set, but given computational limitations we confine ourselves to the five moments mentioned.

---

3 This argument may also explain why the estimator in the paper of Smith and Spencer (1991) does not work well. That estimator uses only three moments, the mean and variance of the exchange rate returns and the variance of the level, to estimate three parameters, so that in principle a unique estimator could be found. However, the variance of the exchange rate level is not very informative about the parameters. This is probably the cause of the difficulties in finding an optimum for the criterion function and obtaining good estimates for the parameters of the target zone model.

3.5 Empirical results.

In this section we present the results of an empirical application of the target zone model to exchange rates of six EMS currencies against the Deutschemark. The model is estimated and it is tested to see whether the model can explain certain "stylised" facts of the exchange rate data.

We use samples of EMS exchange rates from the period after the realignment of January 1987. The sample ends in October 1990, well before the effects of the German unification became known and culminated in the crisis in the EMS in September 1992. Although the absence of realignments does not imply complete credibility of the band, see Svensson (1991c), we think it is worth testing the Krugman model on a sample from a relatively stable period within the EMS. The time series we study are 189 weekly observations from 14 January 1987 until 3 October 1990 of the exchange rates of six major EMS currencies against the Deutschemark. The currencies are the Belgian franc, the Dutch guilder, the Danish krone, the French franc, the Irish punt, and the Italian lira. The dependent variable e_t is defined as 100 times the relative deviation from the central parity. The upper and lower limit on the exchange rate in the model are put equal to the official bound of the target zone, i.e. a deviation of +2.25% upward and -2.25% downward from the central parity is allowed (+6% and -6% for the Italian Lira until January 1990). We use weekly data to avoid problems with missing observations due to weekends and holidays.

The Maximum Likelihood parameter estimates and some associated test statistics are reported in Table 3.1. The estimates of α are significant for three currencies, the Belgian franc, the Danish krone and the French franc. Graphical inspection of the data series shows that these exchange rates come close to the margin, but always revert to the central parity. Also, the volatility of these exchange rates appears to be smaller close to the margin. The target zone model picks up these effects. The large estimates of σ^2 and α for the Danish krone series and their large standard errors are caused by a high correlation between the estimators of these parameters; the estimated asymptotic correlation between $\hat{\sigma}^2$ and $\hat{\alpha}$ is nearly one. The likelihood function is also very flat in these parameters, which is reflected in the large standard errors.

The estimates for the Dutch guilder, the Irish punt and the Italian lira show that the target zone model is essentially linear for these currencies; the point estimate of α is very close to 0, and the hypothesis that $\alpha=0$ cannot be rejected by the Likelihood Ratio test at any usual level of significance. The failure of the target zone model to detect any nonlinearities in the Guilder and Lira rates is probably not too surprising, as the Guilder is always very close to its central parity in the sample, and also the Lira doesn't get close to the margins of its relatively wide target zone, probably due to intra-marginal interventions.

Table 3.1 Maximum Likelihood estimates for Deutschemark exchange rates

currency	μ	σ^2	α	LR	M-norm	M-arch
Belgian franc	-0.148 (0.387)	1.399 (2.124)	4.089 (9.099)	5.32	0.89	1.19
Dutch guilder	0.028 (0.153)	0.170 (0.027)	0.000 (.)	0.00	17.01	170.75
Danish krone	-0.028 (0.017)	4972.03 (4021.01)	139.37 (71.61)	31.47	12.75	1.96
French franc	1.196 (0.710)	8.256 (2.889)	4.375 (1.362)	7.97	316.14	19.92
Irish punt	0.181 (0.710)	1.625 (0.282)	0.008 (0.002)	0.00	8.93	26.47
Italian lira	-1.296 (1.286)	5.723 (1.515)	0.240 (0.443)	0.10	105.18	3.71

Weekly data from 14/01/1987 to 3/10/1990. Asymptotic standard errors in parentheses.
Dependent variable e(t) is $100 \cdot \ln$(exchange rate/central parity).

Unit of parameters μ and σ^2 is 1/year, unit of α is one year.

LR: likelihood ratio test for $\alpha=0 \simeq \chi^2(1)$.

M-norm: M test on third and fourth moment of $\Delta e \simeq \chi^2(2)$.

M-arch: M test on first and second autocorrelation of $\Delta e^2 \simeq \chi^2(2)$.

The Method of Simulated Moments estimates are reported in Table 3.2. The estimates for the guilder and punt are similar to the ML estimates. Note that the estimate of α for these series is close to zero and the model is again essentially linear. In that case, the drift and variance of the fundamental (μ and σ^2) are well estimated by the first two moments of the first difference of the exchange rate. For the other series the MSM estimates are far from the Maximum Likelihood estimates. The estimates of the parameters are imprecise, which is reflected in the large standard errors, but a minimum of the distance function is always found. Thus, our choice of moments improves upon the methods of Smith and Spencer (1991) although precise parameter estimates cannot be obtained. A general specification test of the model is provided by the minimum of the criterion function. If the model is correctly specified, the minimum has a χ^2 distribution with degrees of freedom equal to the number of moments minus the number of parameters. The results in Table 3.2 show that this test rejects the model for the guilder, French franc and lira. Lindberg and Söderlind (1992) argue that this may be due to the misspecification of the Krugman model which does not allow for interventions within the band.

Table 3.2 Simulated moments estimates for Deutschemark exchange rates

currency	μ	σ^2	α	D	M-norm	M-arch
Belgian franc	-0.432 (11.57)	1.505 (77.39)	5.109 (383.66)	2.05	11.99	0.18
Dutch guilder	0.541 (0.248)	0.172 (0.023)	0.100 (0.031)	12.67	15.28	16.23
Danish krone	1.600 (31.40)	9.507 (385.68)	0.867 (44.00)	4.32	4.68	0.19
French franc	1.094 (14.58)	6.853 (182.84)	0.737 (30.75)	47.44	29.30	0.54
Irish punt	0.859 (0.744)	1.719 (0.180)	0.000 (0.003)	1.40	52.73	20.99
Italian lira	1.933 (5.371)	6.278 (5.286)	0.001 (0.082)	84.46	155.76	1.27

Notes: D is the minimum of the distance function $\simeq \chi^2(2)$.
Moments used for estimation: mean, variance, skewness, first and second autocorrelation of the first difference of the exchange rate.
See also notes to Table 3.1.

We also perform two specification tests, which test whether the model can explain certain "stylised" facts about exchange rate returns, in particular ARCH effects and non-normality of the first differences. The test procedure employed is a variant of the M-test developed by Newey (1985). The principle of the test is to check whether some moments of the exchange rate distribution generated by the theoretical model are significantly different from their sample counterparts. The moments used for testing are computed by simulation. Details on the computation of the test statistics are in Appendix B.

The first test, M-norm, is based on the unconditional third and fourth moment of the exchange rate returns. The distributional assumption made in deriving the Krugman model, as in nearly all other target zone models, is that the innovations in the fundamental are normally distributed. The S-shaped transformation from fundamental to exchange rate causes the exchange rate distribution to be non-normal and have even thinner tails than the normal. The M-norm statistic tests whether the observed first differences show a higher degree of departure from normality than is implied by the target zone model.

The second test, M-arch, is based on the autocovariances of the squared exchange rate returns, $(\Delta e_t)^2$. In the target zone model we expect a positive correlation in the second moment of the returns, because due to the S-shaped mapping from fundamental to exchange rate the conditional variance is relatively large when the exchange rate is in the middle of the band, but relatively small close to the bounds. So, if the variance of the fundamentals process is small, successive observations on the exchange rate have a similar position within the band. This implies that the conditional variances of successive returns have similar values, so that there is an endogenous ARCH effect in the target zone model. The question is whether this endogenous ARCH effect in the model is strong enough to explain the observed serial correlation in the second moments of the returns.

The tests in Table 3.1 reveal that for all series except the Belgian franc the model is misspecified. The M-arch test indicates that the ARCH effect in the data on Dutch guilder, French franc and Irish punt rates is not fully explained by the model. However, for the other currencies there is no significant difference in observed and predicted ARCH. The M-norm test on the skewness and kurtosis of the exchange rate returns is highly significant for all series but the Belgian franc. Clearly, the target zone model is not capable of explaining one of the most prominent stylised facts of exchange rates, namely, the fat-tailed distribution of the returns.

3.6 Conclusions.

In this chapter we developed Maximum Likelihood and Method of Simulated Moments estimators for the target zone model of Krugman (1991). It is shown that estimation of this model is feasible. The empirical results indicate that the Krugman model is misspecified for actual EMS exchange rate data. The Maximum Likelihood and Method of Simulated Moments parameter estimates differ substantially. Moreover, the model is not capable to explain two well known stylised facts of exchange rates, autoregressive conditional heteroskedasticity and a non-normal distribution for the returns. The main deficiencies of the Krugman model seem to be the restrictive dynamics, where the scalar fundamental follows a first-order Markov process, the normal distribution of this process, and the assumption that there are no interventions within the band.

Appendix A. Simulating the target zone model.

The method of simulated moments estimator and the M tests discussed in this chapter require simulation of the stochastic process of the target zone model. In this appendix we describe how the numerical simulations are performed. The method is based on the work of Duffie and Singleton (1988) who developed a method for computing the moments of a Brownian motion. The difficulty of a discrete time computer simulation of the continuous time regulated Brownian motion is caused by the assumption that there is *infinitesimal* intervention at the margins, so that the process does not jump and has a zero probability of being exactly at the margin. In a discrete time approximation, interventions are strictly positive, and we have to make some assumption on where the process goes after an intervention. The scheme used by Smith and Spencer (1991) and Beetsma (1991) is

$$f^*(t+\Delta t) = f(t) + \mu\Delta t + \sigma\sqrt{\Delta t}\cdot\hat{\varepsilon}(t)$$

$$f(t+\Delta t) = \begin{cases} \bar{f} \text{ if } & f^*(t+\Delta t) > \bar{f} \\ \underline{f} \text{ if } & f^*(t+\Delta t) < \underline{f} \\ f^*(t+\Delta t) & \text{otherwise} \end{cases}$$

It is clear that for $\Delta t>0$ there will be a point mass at $\bar{f}$ and $\underline{f}$ in the distribution of $f(t+\Delta t)$, whereas the mass at those points in the continuous time model is zero. This point mass can be large if $f(t)$ is close to $\bar{f}$ or $\underline{f}$ relative to the magnitude of Δt. A way to improve the accuracy of the simulation is to choose Δt very small if $f(t)$ is close to $\bar{f}$ or $\underline{f}$.

A scheme that gives no point mass at the bounds is found by reflecting the stochastic process in the upper or lower bound if $f^*(t+\Delta t)$ exceeds that bound:

$$f^*(t+\Delta t) = f(t) + \mu\Delta t + \sigma\sqrt{\Delta t}\cdot\hat{\varepsilon}(t)$$

$$f(t+\Delta t) = \begin{cases} 2\bar{f}-f^*(t+\Delta t)-f(t) & \text{if } f^*(t+\Delta t)>\bar{f} \\ 2\underline{f}-f^*(t+\Delta t)-f(t) & \text{if } f^*(t+\Delta t)<\underline{f} \\ f^*(t+\Delta t) \text{ otherwise} \end{cases}$$

We prefer this scheme because it generates no observations exactly on the bounds. Such observations cause problems in computing the likelihood function for simulated data because, as a result of the smooth pasting conditions, $G'(f)$ for such observations is zero, so that the Jacobian of the transformation from fundamental to exchange rate and hence the likelihood function are infinite.

The conditional distribution function of $f(t+\Delta t)$ generated by this scheme with initial value $f(t)=f_t$ is

$$P(f|f_t) = \Phi\left(\frac{f-f_t-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right) - \Phi\left(\frac{2\underline{f}-f-f_t-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right) + (1-\Phi\left(\frac{2\bar{f}-f-f_t-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right))$$

The first part is the usual normal distribution function, the second part represents the probability that f(t) has been reflected in the lower bound, and the third part is the probability mass reflected in the upper bound. Comparing this distribution with the approximate distribution function of the continuous time model, (3.13), we see that the discrete simulation overestimates the probabilities of reflection somewhat, due to the omission of interventions that possibly take place within the time interval $(t,t+\Delta t]$. So, for a good approximation to the continuous time distribution it is necessary to use a simulation interval Δt that is small compared with the length of the observation interval. In our applications we use 10 drawings per observation.

Appendix B. Tests of moment restrictions.

In this chapter several tests of moment restrictions of the form $Ld(\theta)=0$ appear. In this appendix we discuss how to test these restrictions by Newey's (1985) M-test. Typically, $d(\theta)$ is estimated by $d_T(\theta_T)$, the distance between observed sample moments and the moments simulated from the target zone model, using a consistent estimate θ_T of the parameter vector as the pseudo-true value. Let Q be the asymptotic covariance matrix of $\sqrt{T}\cdot d_T(\theta_T)$, and let $\hat{Q}_T$ be a consistent estimator of Q. Under the null, the statistic

$$M = T \cdot d_T(\theta_T)' L' (L Q_T L')^{-1} L d_T(\theta_T)$$

has an asymptotic chi-square distribution with degrees of freedom equal to the rank of LQL'. This rank is usually equal to the dimension of $Ld_T(\theta_T)$, unless some moments were used to obtain the estimator θ_T or some moments are collinear. The difficult part of computing the test is finding the right expression for Q.

Case 1. Maximum Likelihood estimator.

Denote the gradient of the log-likelihood function by $g_T(\theta)$ and the distance vector by $d_T(\theta)$. The covariance matrix of $(g_T(\theta)', d_T(\theta)')'$ is partitioned as follows

$$V = \begin{bmatrix} V_{gg} & V_{gd} \\ V_{dg} & V_{dd} \end{bmatrix}$$

Newey (1985) proves that the asymptotic distribution of the moments vector, evaluated in the estimate θ_T is

$$\sqrt{T} \cdot d_T(\theta_T) \underset{asy}{\sim} N\left(0,\ V_{dd} - V_{dg} V_{gg}^{-1} V_{gd}\right)$$

Intuitively, the variance of $d_T(\theta_T)$ is equal to the variance of the moments evaluated in the true value plus a correction for the fact that the moments are evaluated in the estimated parameter vector.

Case 2. Method of moments estimator.

Let $d_T(\theta)$ now denote the vector of all moments used for estimation and testing. Suppose the moments used for estimation are picked from the vector of all moments by the selection matrix S and an estimated optimal weight matrix is used. The variance-covariance matrix of the estimator then is

$$V(\theta_T) = (FD)^{-1}, \quad D \equiv \partial d(\theta_0)/\partial\theta, \quad F \equiv (SD)'(SV_{dd}^{-1}S')^{-1}S$$

Newey proves that in this case, the asymptotic variance of $d_T(\theta_T)$ is

$$\sqrt{T} d_T(\theta_T) \underset{asy}{\sim} N\left(0,\ V_{dd} - D(FD)^{-1}FV_{dd} - V_{dd}F'(D'F')^{-1}D' + DV(\theta_T)D'\right)$$

Chapter 4

Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates

4.1 Introduction.

The recent literature on modelling exchange rates in a target zone relies heavily on the use of continuous time stochastic models and stochastic calculus. Although this approach is neat and leads to interesting theoretical insights, the empirical applicability is limited and the fit of these models to the data is generally poor. One of the likely causes of this poor fit is the almost exclusive reliance on normal distributions in these models. Exchange rates exhibit properties, such as a fat tailed return distribution and strong conditional heteroskedasticity, that are not easily incorporated in Gaussian continuous time models.

In this chapter we propose a discrete time alternative to the continuous time target zone models. In the target zone the exchange rate is kept within a band by both marginal and intra-marginal interventions. The approach followed has several advantages over the usual models. First, once the discrete time model is solved, the likelihood function and moments of the distribution of the discretely observed data are available immediately; there is no need to compute complex transition densities like we did in Chapter 3. Second, more general non-Gaussian stochastic processes are easily incorporated in the discrete time framework.

In the empirical part, the model is estimated on Deutschemark exchange rates from the EMS. Using the estimates of the parameters of the model, time series of expected depreciation within the band are constructed. These series are used to assess the credibility of the EMS target zones by the method of Svensson (1991d). In this method, the interest rate differential with the central country is corrected for the expected depreciation of the currency within the band. The remaining part reflects the expected devaluation due to realignments of the central parity.

The setup of the chapter is as follows. In Section 2, we present the discrete time model and provide a method to solve the model numerically. In Section 3, the model is estimated on EMS exchange rate data assuming a fully credible target zone. In Section 4 we assess the credibility of the EMS target zones by comparing predicted depreciations within the band with observed interest rate differentials. In Section 5 we conclude.

4.2 The model.

The basis of the model is the discrete time equivalent of (3.1),

$$(4.1) \qquad e_t = f_t + \lambda E_t e_{t+1}, \quad \lambda = \alpha/(1+\alpha), \quad \alpha \geq 0,$$

where e_t denotes the spot exchange rate, f_t the "fundamental" and E_t the conditional expectation given all information available at time t. Solving the model forward and excluding bubbles one obtains the "present value" solution of the exchange rate

$$(4.2) \qquad e_t = \sum_{i=0}^{\infty} \lambda^i E_t f_{t+i},$$

In a discrete time framework, the exchange rate under a target zone regime is a limited dependent variable on the interval $[\underline{e},\bar{e}]$. There are different ways to model this property. Pesaran and Samiei (1992) start from equation (4.1) and "censor" the exchange rate at the upper and lower boundary, so that the model becomes

$$(4.3) \qquad e_t = C(f_t + \lambda E_t e_{t+1}; \underline{e},\bar{e}),$$

where C(.) is the censoring function that is defined as follows:

$$(4.4) \qquad C(x; \underline{e},\bar{e}) = \begin{cases} \underline{e} & \text{if } x < \underline{e} \\ x & \text{if } \underline{e} < x < \bar{e}. \\ \bar{e} & \text{if } x > \bar{e} \end{cases}$$

Pesaran and Samiei (1992) assume that the fundamental is a linear combination of observed variables, generated by a VAR process. In line with most current research we do not assume that the fundamental is a function of observed economic variables. Instead, we assume that f_t is unobserved and follows a mean reverting Markov process that is censored to the interval $[\underline{f},\bar{f}]$:[1]

$$f_{t+1} = C((1-\varphi)\mu + \varphi f_t + \varepsilon_{t+1}; \underline{f},\bar{f}). \tag{4.5}$$

As a consequence of the censoring of the fundamental, the exchange rate is also bounded and thus fluctuates in a target zone. This property is intuitively clear from (4.2), where due to the censoring the support of all future fundamentals is bounded.

The parameter φ is the degree of mean reversion in the fundamental. If $\varphi=1$, the fundamental follows a random walk, which is the discrete time equivalent of the continuous time Brownian Motion assumed by Krugman (1991). If $\varphi<1$, there is mean reversion of the fundamental, possibly caused by intra-marginal interventions of the central banks. This model corresponds to the continuous time Ornstein-Uhlenbeck process used by Lindberg and Söderlind (1992) to model Swedish exchange rates. Note that in (4.5) the mean, μ, disappears if the fundamental is a random walk ($\varphi=1$). The reason for this parametrisation is that we want to exclude deterministic trends in the fundamental. The bounds on the fundamental are restricted to be symmetric around zero, i.e. $\underline{f}=-\bar{f}$, because otherwise μ is not identified.

The discrete time model allows a general flexibility for the distribution of the innovations in the fundamental. Whereas the continuous time models exclusively rely on normal distributions (Wiener processes), in the discrete time model any known density can be used. In particular, distributions with fatter tails than the normal are likely to empirically dominate the normal and are easily implemented in the model. For the solution method presented here, it is necessary that the shape of the distribution function does not depend on time or on previous state variables. This excludes for example the ARCH process where the conditional variance is a function of the previous innovations.

We now turn to the solution of the model given by (4.1) and (4.5). Because of the censoring the model is non-linear and cannot be solved by standard methods designed for linear Rational Expectations models[2]. However, the model has a

---

[1] This model is an extension of the model of Koedijk, Stork and De Vries (1993) who assume that the fundamental is a censored random walk.

[2] Pesaran and Samiei (1992b) propose to define an intervention variable which equals the 'latent' exchange rate from (4.1) minus the actual. Then the infinite summation (4.2) is then rewritten to include all expected future interventions, and the exchange rate is computed by backward recursion. The problem with this

structure that is similar to the problems of commodity pricing with speculative storage analysed by Deaton and Laroque (1991a) and the model of saving with liquidity constraints analysed by Deaton (1991). Deaton and Laroque (1991a,b) have developed methods to solve and estimate these models, and their methods are easily adapted to solve the discrete time target zone model.

The first thing to note is that the Markov property of the fundamental implies that the current value of the fundamental is a sufficient statistic for the distribution of future fundamentals. As a consequence, the expected future values in (4.2) and hence the exchange rate e_t are a function of the current fundamental f_t only. Therefore, it is possible to define an exchange rate function that gives the exchange rate for every value of the fundamental, $e=G(f)$. This function should satisfy the following functional equation, derived from (4.1) and (4.5):

$$(4.6) \qquad G(f) = f + \lambda\int G(v)dP_{\varepsilon}(v|f), \qquad v = C((1-\varphi)\mu + \varphi f + \varepsilon; \underline{f},\bar{f}),$$

where $P_{\varepsilon}(v|f)$ denotes the distribution of v given $f_t=f$. The integral on the right hand side corresponds to the expected exchange rate in the next period, given the fundamental in the current period.

We closely follow the approach outlined in Deaton and Laroque (1991b) to solve this equation. The function G(f) is approximated on a grid of points $f_1,..,f_N$ on the relevant range of the fundamental. If the bounds on the fundamental were known, this range would of course be $[\underline{f},\bar{f}]$. In practice, the band on the exchange rate is known and $\underline{f}$ and $\bar{f}$ must be solved from the 'value matching' conditions, $\underline{e}=G(\underline{f})$ and $\bar{e}=G(\bar{f})$. In our model, the exchange rate solution is always bounded between the 45° line $e=f$ and the free float solution $e=f/(1-\lambda)$. Hence, the relevant range of the fundamental is always contained in the target zone interval of the exchange rate, $[\bar{e},\underline{e}]$. Therefore, we use an equidistant grid $f_1,..,f_N$ on $[\bar{e},\underline{e}]$. In the actual computations, 50 points were sufficient.

Having defined the grid, the next step is to compute the function values $g_i=G(f_i)$ on all points in the grid. This is not feasible in one step, because the expected value of the exchange rate, E_te_{t+1}, depends on the function G(f). The way to solve this problem is to start from an initial guess of the function values, $g_1^{(0)},..,g_N^{(0)}$, on the grid, feed these into the expectation formula, and compute new function values. This method will work if (4.6) is a contraction mapping, see

approach, however, is that the interventions at time $t+\tau$ ($\tau>0$) depend on the expected interventions at $t+\tau+1$, so that the recursion becomes quite complicated.

Sargent (1987). In the Appendix we prove that (4.6) is indeed a contraction mapping, so this method will always converge to the right solution.

Numerical computation of the integral in (4.6) goes as follows. First, split the range of f in three parts, $f \leq \underline{f}$, $\underline{f} < f < \bar{f}$ and $f \geq \bar{f}$:

$$(4.7) \qquad G(f) = f + \lambda\left(P_\varepsilon(\underline{f}\,|\,f)\underline{e} + [1-P_\varepsilon(\bar{f}\,|\,f)]\bar{e} + \int_{\underline{f}}^{\bar{f}} G(v)p_\varepsilon(v\,|\,f)dv\right)$$

The latter integral can be approximated by a weighted summation over the initial function values, where the weights are given by the conditional densities of the fundamental at the j^{th} grid point, $p_\varepsilon(f_j\,|\,f_i)$:

$$(4.8) \qquad g_i^{(n)} = f_i + \lambda\left(P_\varepsilon(\underline{f}\,|\,f_i)\underline{e} + [1-P_\varepsilon(\bar{f}\,|\,f_i)]\bar{e} + \sum_{\underline{f}<f_j<\bar{f}} g_j^{(n-1)}p_\varepsilon(f_j\,|\,f_i)\right)$$

A good choice for the initial solution turned out to be the censored free float solution for a random walk fundamental: $g_i^{(0)}=C(f_i/(1-\lambda);\underline{e},\bar{e})$.

Figure 4.1 shows an example of the solution to the model where the fundamental is a driftless random walk ($\varphi=1$) with normally distributed innovations with variance $\sigma^2=0.003$. The parameter of the expected exchange rate λ is chosen to be 0.8. These parameter values are roughly equal to the estimated values for the Danish kroner-Deutschemark exchange rate. In the figure the exchange rate solution $e=G(f)$ is plotted against the free float solution $f/(1-\lambda)$. The solution shows the familiar S-shape. However, there is no smooth pasting on the edges of the band.

Further insight in the properties of the discrete time target zone model can be gained from the conditional expectation and conditional variance of the exchange rate one period ahead. Generally, conditional moments of the exchange rate, $E[h(e_{t+1})\,|\,e_t]$, are easily computed by integrating the function $h(G(f_{t+1}))$ over $P_\varepsilon(f_{t+1}\,|\,f_t)$ given $f_t=G^{-1}(e_t)$:

$$(4.9) \qquad E[h(e_{t+1})\,|\,e_t] = \int h(G(v))dP_\varepsilon(v\,|\,f_t), \quad v = C((1-\varphi)\mu + \varphi f_t + \varepsilon;\ \underline{f},\bar{f}),$$

which is numerically approximated by the appropriate modification of (4.7). The analysis of the same random walk specification as before reveals some interesting properties of the model. First, there is endogenous mean reversion of exchange rates in the model. This is illustrated in Figure 4.2, where the expected exchange rate (solid line) is plotted against the current exchange rate (dotted line). If the exchange rate were a random walk, the current and expected exchange rate would be the same and the graph would be equal to the 45° line. In figure 4.2, the expected exchange rate is always smaller than the current if e_t is above the central parity. This mean reversion within the band is generated endogenously in the target zone model. As is clear from the figure, the endogenous mean reversion is not strong. Empirically it is probably necessary to include mean reversion in the fundamental ($\varphi < 1$) as well.

Figure 4.3 shows the conditional variance of the exchange rate. There is endogenous conditional heteroskedasticity in the model because the conditional variance is lower close to the bounds and higher in the middle because of the possible censoring in the next period. A second thing to note is that the conditional variance of the exchange rate is smaller than the conditional variance of the fundamental. This is because the S-shaped relation between fundamental and exchange rate reduces the variance. This is Krugman's famous honeymoon effect.

Figure 4.1: Discrete time target zone solution

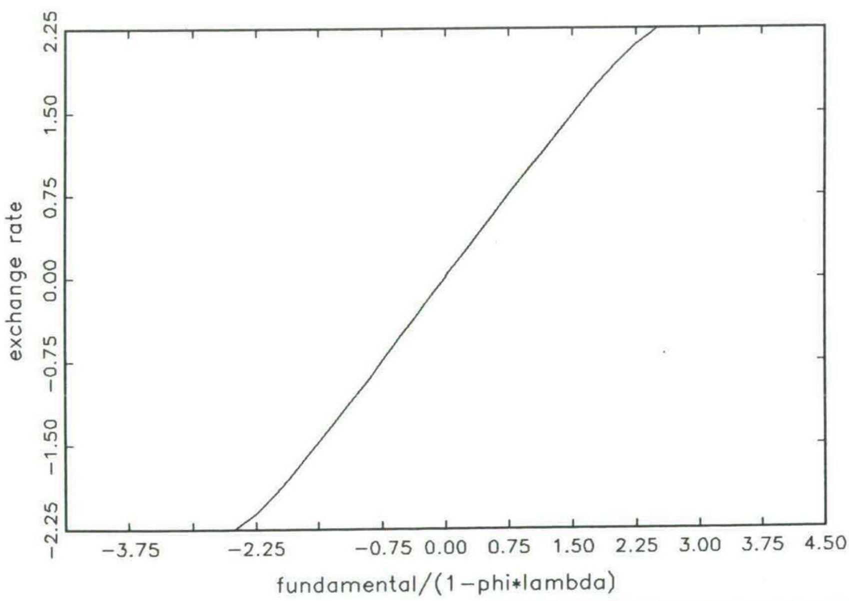


Figure 4.2: Conditional mean of exchange rate

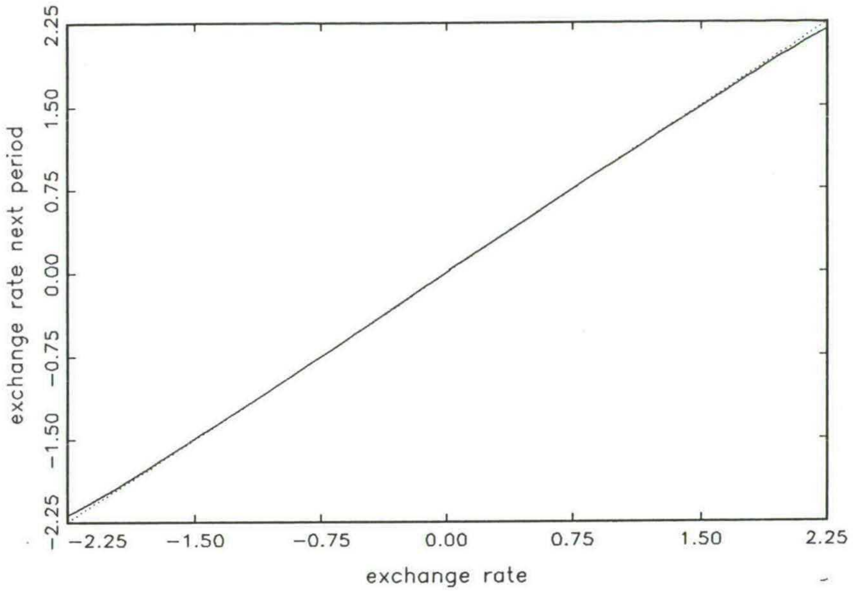
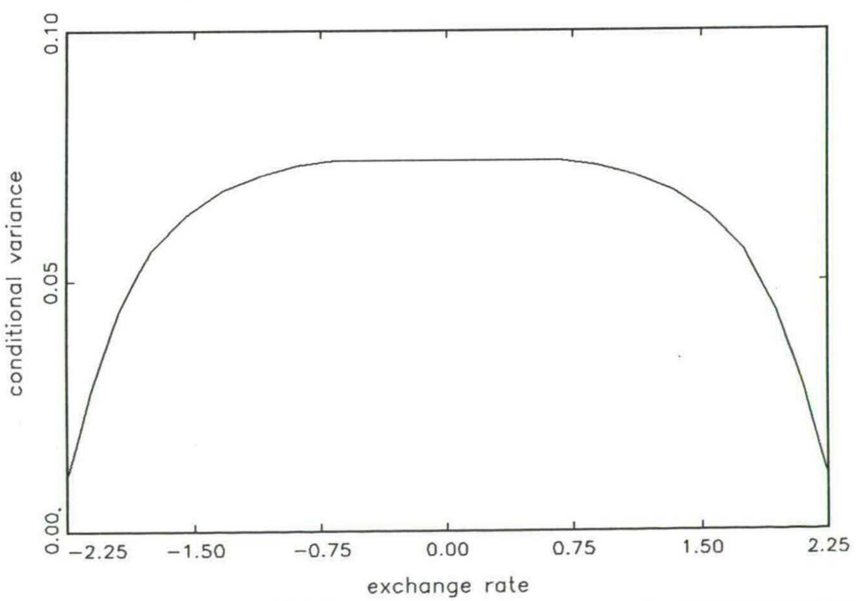


Figure 4.3: Conditional variance of exchange rate



Summarising, the discrete time target zone model has all the properties of the continuous time models, including the mitigating effect of the S-shaped exchange rate function. The advantage of the discrete time model lies in its flexibility to include different distributions for the fundamental. Which distribution is the best is an empirical question that is addressed in the next section.

4.3 Estimating the discrete time target zone model.

Estimates of the model can be obtained in several ways. A simple but inefficient method is to select some moments of the *marginal* (unconditional) distribution of the exchange rate and estimate the model by GMM. Because of the truncation of the fundamental in every period, the marginal moments must be computed by simulation. In Chapter 3 we argued that this method is inefficient compared with likelihood-based estimators as well as computationally unattractive. There are several alternative estimators. Deaton and Laroque (1991b) suggest to use the Quasi Maximum Likelihood estimator of Gourieroux et al. (1984), which is based on the first two conditional moments. We estimate the model efficiently by Full Information Maximum Likelihood[3].

The likelihood function is obtained as an immediate by-product of the solution to the model obtained in the previous section. Given a series of observations on the exchange rate $(e_1,..,e_T)$ and conditional on the first observation e_0, the likelihood function is equal to the sum of the log-likelihood contributions, ℓ_t, of each observation:

$$\ln L(\theta) \equiv \sum_{t=1}^{T} \ell_t \tag{4.10}$$

Because the exchange rate is a function of the current fundamental only, and because the fundamental is a first order Markov process, the exchange rate is a Markov process as well. As explained in Chapter 3, by the prediction error decomposition the likelihood contributions are derived from the conditional distribution of e_t given e_{t-1}. The likelihood function is of the two-limit Tobit type. The likelihood contribution for an observation at the one of the bounds is equal to the probability that the exchange rate hits that bound; for an observation within the band, the likelihood contribution is equal to its conditional density:

---

3 Deaton and Laroque also mention the FIML estimator, but they run into problems because their model has a discontinuity, which causes non-differentiability of the likelihood function. This is not a problem in the discrete time target zone model.

$$(4.11a)\qquad \ell_t = \ln P_\eta(\underline{e}|e_{t-1}) \qquad \text{if } e_t=\underline{e}$$

$$(4.11b)\qquad \ell_t = \ln p_\eta(e_t|e_{t-1}) \qquad \text{if } \underline{e}<e_t<\bar{e}$$

$$(4.11c)\qquad \ell_t = \ln(1-P_\eta(\bar{e}|e_{t-1})) \qquad \text{if } e_t=\bar{e}$$

where $P_\eta(.|e_{t-1})$ denotes the conditional distribution of the current exchange rate, e_t, given the previous exchange rate, e_{t-1}.

The conditional distribution of the exchange rate is only implicit in the model, and a transformation to the latent fundamental is needed. The fundamental is constructed from the observed exchange rates by solving the equation $e_t=G(f_t)$ for f_t, given the value of the parameter vector θ. The observations of the exchange rate are generally different from the points in the grid $(g_1,..,g_N)$, so that the exact value of the fundamental is not immediately available. Therefore, the function G(f) is approximated by cubic spline interpolation[4]. This method has the advantage that the approximation is twice continuously differentiable and easily inverted.

Having computed the fundamentals, the likelihood contributions are obtained by a the change-of-variables technique described in Chapter 3. For observations at the bounds, the contribution is equal to the probability that the fundamental hits the band[5]. The likelihood contribution of an observation inside the band is slightly more complicated, because it involves the Jacobian of the transformation from exchange rate to fundamental. After the transformation, the likelihood contributions are as follows:

$$(4.12a)\qquad \ell_t = \ln P_\varepsilon(\underline{f}|f_{t-1}) \qquad \text{if } e_t=\underline{e}$$

$$(4.12b)\qquad \ell_t = \ln p_\varepsilon(f_t|f_{t-1}) - \ln G'(f_t) \qquad \text{if } \underline{e}<e_t<\bar{e}$$

$$(4.12c)\qquad \ell_t = \ln(1-P_\varepsilon(\bar{f}|f_{t-1})) \qquad \text{if } e_t=\bar{e}$$

4 We used the algorithm described in Cheney and Kincaid (1985, p.275-6).

5 Recall that in the continuous time model, the probability of having an observation on the band is zero, because of the continuous marginal intervention. In the discrete time model, however, this probability is non-zero.

where P_ε denotes the conditional distribution of the fundamental, which is known by assumption. The second term of (4.12b) is the logarithm of the Jacobian of the transformation from e_t to f_t.

Given a choice for the conditional distribution of the fundamental, the likelihood function is complete. In the empirical work we use two distributions, the normal and the Student-*t* distribution. For the autoregressive model the associated density functions are

Normal distribution

$$p_\varepsilon(f_t|f_{t-1}) = (2\pi\sigma^2)^{-1/2}\exp\left(\frac{(f_t-(1-\varphi)\mu-\varphi f_{t-1})^2}{\sigma^2}\right)$$

Student-t distribution

$$p_\varepsilon(f_t|f_{t-1}) = \frac{\Gamma((\nu+1)/2))}{\Gamma(\nu/2)\,\Gamma(1/2)}(\nu\sigma^2)^{-1/2}\left(1 + \frac{(f_t-(1-\varphi)\mu-\varphi f_{t-1})^2}{\sigma^2}\right)^{-\frac{(\nu+1)}{2}}$$

where Γ denotes the Gamma function and ν is the degrees of freedom parameter of the t distribution. The *t* distribution reduces to the normal for $1/\nu=0$.

Maximization of the log-likelihood function with respect to the parameters σ^2, φ and λ yields a consistent, efficient and asymptotically normal estimator. If one uses the Student-*t* distribution, $1/\nu$ is added as an extra parameter to be estimated. The likelihood function is continuous in the parameters because the exchange rate solution G(f) is continuously differentiable. If cubic spline interpolation is used the numerical approximation to G(f) is (twice) continuously differentiable as well, so that standard optimisation routines can be applied. Asymptotic standard errors are obtained by inverting the outer product of gradient matrix.

We now turn to the empirical results on EMS data. Initially, the most recent EMS period without realignments is chosen. Of course, absence of realignments does not imply that the target zone is fully credible. We shall therefore test the credibility of the target zone afterwards. The data used are weekly spot exchange rates from 14 January 1987 to 23 September 1992 [6]. This sample is slightly longer

[6] These data were kindly provided by Philip Stork.

than the sample used in Chapter 3. As before, the exchange rate is expressed as the percentage deviation from the central parity.

Table 4.1 reports the empirical results for the discrete time target zone model given by (4.1) and (4.5), under the assumption that the mean $\mu=0$. The innovations in the fundamental follow a *t*-distribution.

Table 4.1 Estimates of the fully credible discrete time target zone model on Deutschemark exchange rates; short sample (14/1/1987-23/9/92).

currency	σ^2	φ-1	λ	ν	lnL(θ)	LR_ν	LR_λ
Belgian franc	0.00388 (0.01269)	-0.037 (0.015)	0.741 (0.471)	5.024 (1.471)	-25.85	25.40	1.38
Dutch guilder	0.00231 (0.00432)	-0.165 (0.030)	0.415 (0.672)	4.370 (1.347)	282.21	27.64	0.00
Danish krone	0.00296 (0.00119)	-0.024 (0.014)	0.805 (0.046)	7.046 (3.019)	-45.68	9.40	2.98
French franc	0.00174 (0.00236)	-0.013 (0.010)	0.744 (0.177)	4.613 (1.156)	62.80	39.82	1.88

Standard errors in parentheses; LR_ν is likelihood ratio statistic against normal distribution ($1/\nu=0$). LR_λ is likelihood ratio statistic against $\lambda=0$. Restriction $\mu=0$ imposed.

The table shows several interesting results. For all series, the mean reversion within the band is significant and φ is different from 1 at 5% level [7]. Another important result is that the normal distribution is not an adequate description of the fundamental; the likelihood ratio test of a normal distribution against a t distribution is always way above the $\chi^2(1)$ critical values at usual significance levels. The parameter λ that determines the forward looking aspect of the model is never significant at 5% level, and for the Danish krone series only at 10% level. Thus, the non-linear effect of the target zone is not very pronounced, which may be due to a lack of credibility of the band. We address that issue in the next section.

In Table 4.2 the estimates of the model with $\lambda=0$ are presented; the restriction that $\mu=0$ is now dropped. In this case, the exchange rate function G(f) is linear, $e=f/(1-\lambda)$, and the model reduces to a standard two-limit Tobit model.

[7] Both under the null ($\varphi=1$) and the alternative ($\varphi<1$) the exchange rate is stationary in this model. Therefore, the usual critical values for the t-test apply and not the Dickey-Fuller critical values.

Compared with the model with $\lambda > 0$, the results are rather similar, except that the standard errors are much smaller.

Table 4.2 Estimates of the fully credible discrete time target zone model on Deutschemark **exchange rates, $\lambda = 0$. Short sample (14/1/1987 - 23/9/92)**

currency	μ	σ^2	φ-1	ν	lnL(θ)
Belgian franc	0.577 (0.257)	0.04572 (0.00664)	-0.056 (0.015)	4.865 (0.096)	-25.10
Dutch guilder	-0.010 (0.031)	0.00544 (0.00085)	-0.163 (0.030)	4.379 (0.115)	282.49
Danish krone	0.001 (0.832)	0.06080 (0.00856)	-0.024 (0.018)	7.565 (0.083)	-46.43
French franc	0.000 (1.162)	0.02432 (0.00320)	-0.014 (0.016)	4.540 (0.090)	61.83

Standard errors in parentheses. $\lambda=0$ restriction.

Table 4.3 presents estimates of the model on a longer sample, that covers the complete EMS period. Observations just after realignments are excluded from the analysis by setting their log-likelihood contribution to zero. The estimates over the longer sample differ in two respects from the estimates over the short sample: both the kurtosis ($1/\nu$) and the variance of the fundamentals distribution are higher. Probably this reflects the greater uncertainty in the early EMS which showed numerous realignments.

Table 4.3 Estimates of the fully credible discrete time target zone model on Deutschemark **exchange rates, $\lambda = 0$. Long sample (13/4/79 - 23/9/92)**

currency	μ	σ^2	φ-1	ν	lnL(θ)
Belgian franc	0.863 (0.181)	0.06090 (0.00520)	-0.064 (0.013)	3.373 (0.055)	-241.81
Dutch guilder	-0.118 (0.181)	0.01359 (0.00138)	-0.031 (0.010)	2.331 (0.145)	178.63
Danish krone	0.030 (0.809)	0.05516 (0.00475)	-0.011 (0.005)	3.653 (0.043)	-201.94
French franc	0.007 (1.162)	0.03577 (0.00310)	0.005 (0.005)	3.394 (0.057)	-64.94

Standard errors in parentheses. $\lambda=0$ restriction. Weeks with realignments excluded from sample.

4.4 Assessing target zone credibility.

Svensson (1991d) developed a method to assess the credibility of a target zone. Assuming that Uncovered Interest Parity (UIP) holds, the interest rate differential of a currency versus (here) the Deutschemark is equal to the expected rate of depreciation of that currency vis à vis the Mark over the relevant holding period. The expected rate of depreciation consists of two parts. The first part is the expected rate of depreciation within the band. Svensson argues that the annualised expected rate of depreciation within the band can be fairly large, so that large short term interest rate differentials need not imply a high risk premium or an expected realignment. The second part of the expected rate of depreciation is what Svensson calls the *expected rate of devaluation*, which is defined as the expected change in the central parity due to realignments, corrected for the change of the position of the exchange rate within the band. Bertola and Svensson (1993) argue that the latter correction is not trivial, because with realignments the exchange rate usually jumps from a weak position in the old band to a strong position in the new band.

The decomposition can be formalised as follows (see also Beetsma (1992)). Let s_t denote the spot rate, c_t the central parity, e_t the spot rate "within the band", i.e. the spot rate minus the central parity, and p_t the probability of a realignment. Let $E[.|R]$ denote the conditional expectation, given a realignment, and $E[.|NR]$ the conditional expectation given no realignments. Then the expected rate of depreciation can be decomposed as

$$
\begin{aligned}
E_t[\Delta s_{t+1}] &= E_t[\Delta c_{t+1}] + E_t[\Delta e_{t+1}] \qquad (4.13)\\
&= p_t(E_t[\Delta c_{t+1}|R] + E_t[\Delta e_{t+1}|R]) + (1-p_t)E_t[\Delta e_{t+1}|NR]\\
&= p_t(E_t[\Delta c_{t+1}|R] + E_t[\Delta e_{t+1}|R] - E_t[\Delta e_{t+1}|NR]) + E_t[\Delta e_{t+1}|NR]\\
&\equiv E_t[d_{t+1}] + E_t[\Delta e_{t+1}|NR],
\end{aligned}
$$

where $E_t[d_{t+1}]$ denotes the expected rate of devaluation. Assuming UIP, the (annualised) unconditional expected rate of depreciation is equal to the interest rate differential

$$(4.14) \qquad E_t[\Delta s_{t+1}]/\tau = i_t - i_t^*$$

where i_t is the one-period domestic interest rate, i_t^* the one-period foreign interest rate and τ the length of one period in years. Assuming that the estimated model is correct, the expected future exchange rate within the band is easily calculated from equation (4.9). We use the Maximum Likelihood estimates of the model's parameters in Table 4.2 to compute the expected exchange rates. This is different from Rose and Svensson (1991) and Beetsma (1992) who use a cubic function to approximate the conditional expectation of the exchange rate within the band.

Having computed the expected exchange rate within the band, $\hat{E}_t[e_{t+1}|NR]$, we can construct the expected rate of devaluation using data on the interest rate differential as follows:

$$\text{(4.15)} \qquad \hat{E}_t[d_{t+1}] = \tau(i_t - i_t^*) - \hat{E}_t[\Delta e_{t+1}|NR].$$

Figures 4.4A-D show the spot exchange rate within the band, the interest rate differential, the expected rate of depreciation within the band, and the expected rate of devaluation of four EMS currencies vis à vis the Deutschemark over the short sample period (87/1/14-92/9/23). All graphs show that even in the relatively stable years of the EMS between 1987 and 1992 there have been considerable expected devaluations in all series, although the credibility of the band increased over the years. We confirm the result of Rose and Svensson (1991) who estimated that the probability of realignment for the Danish krone was rather high in 1989, and larger than suggested by the interest rate differential only.

Another interesting results concerns the estimated rate of devaluation of the Belgian franc. The estimates suggest that the expected rate of devaluation became very small or even negative from week 160 (early 1990) onwards, whereas the interest rate differential disappeared much later, from early 1991 onward. Thus we cannot ignore the expected rate of depreciation within the band if we want to assess target zone credibility.

Figure 4.4A Data and expected rate of realignment for Belgian franc / Deutschemark exchange rate.

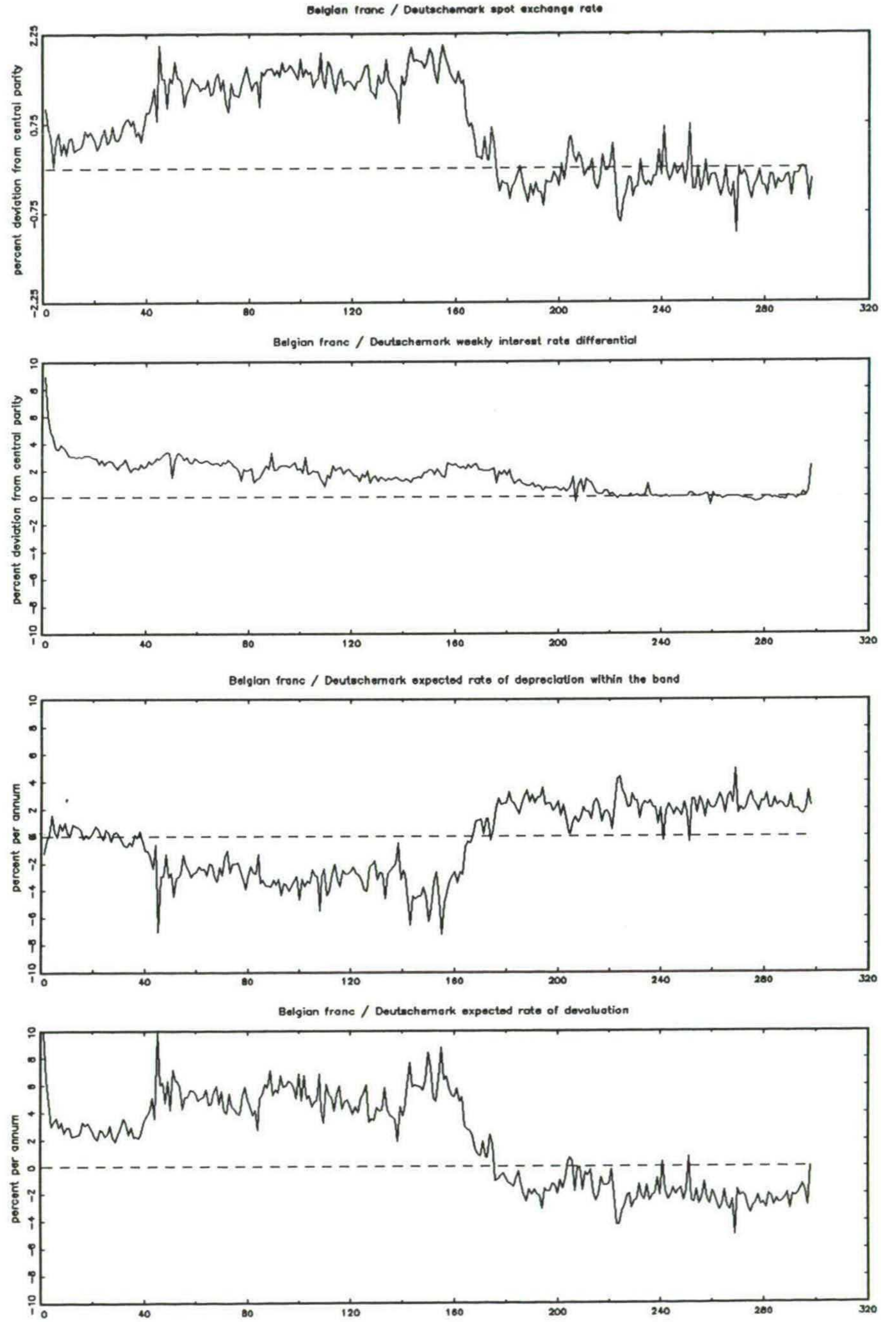


Figure 4.4B Data and expected rate of realignment for Dutch guilder / Deutschemark exchange rate.

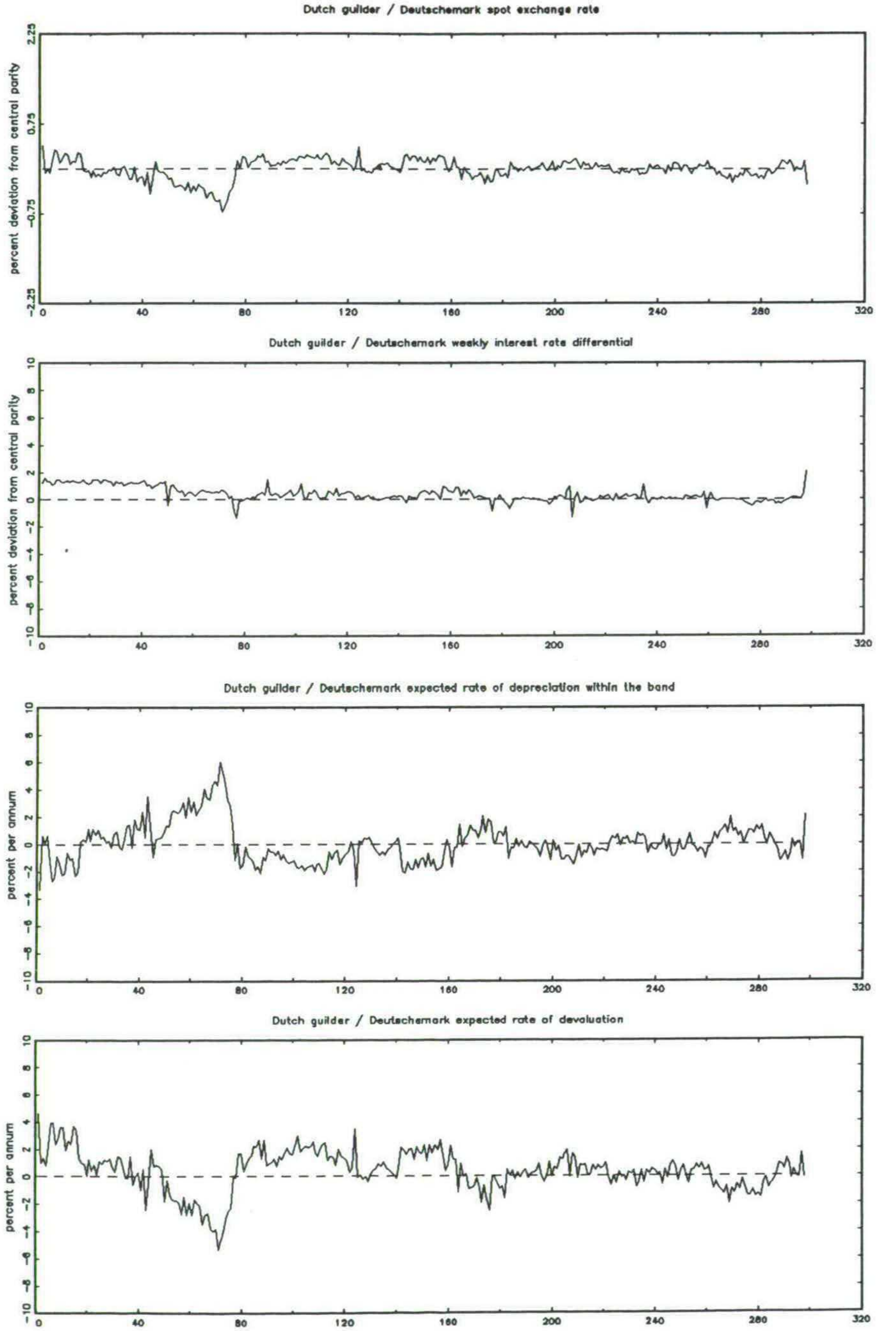


Figure 4.4C Data and expected rate of realignment for Danish krone / Deutschemark exchange rate.

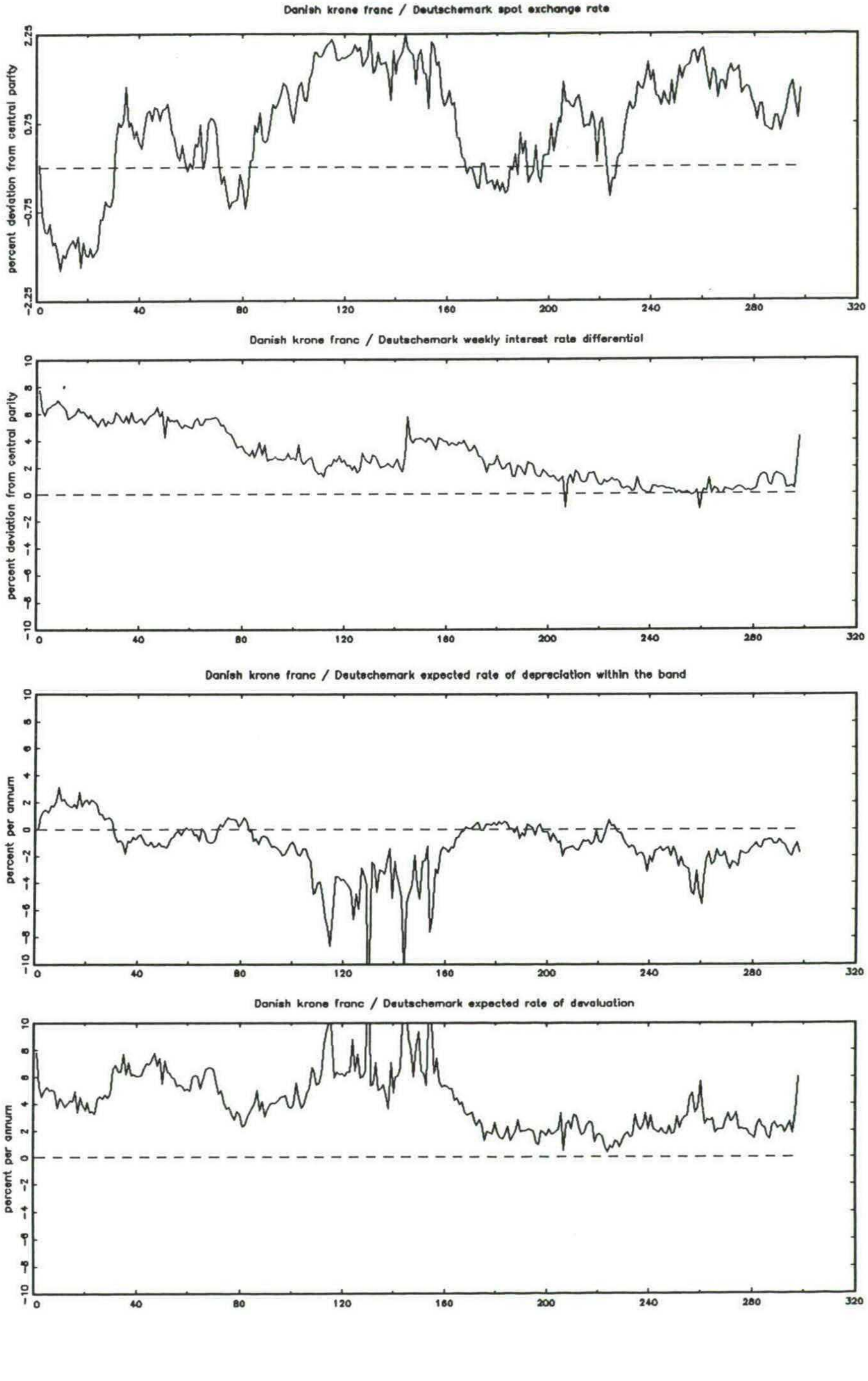
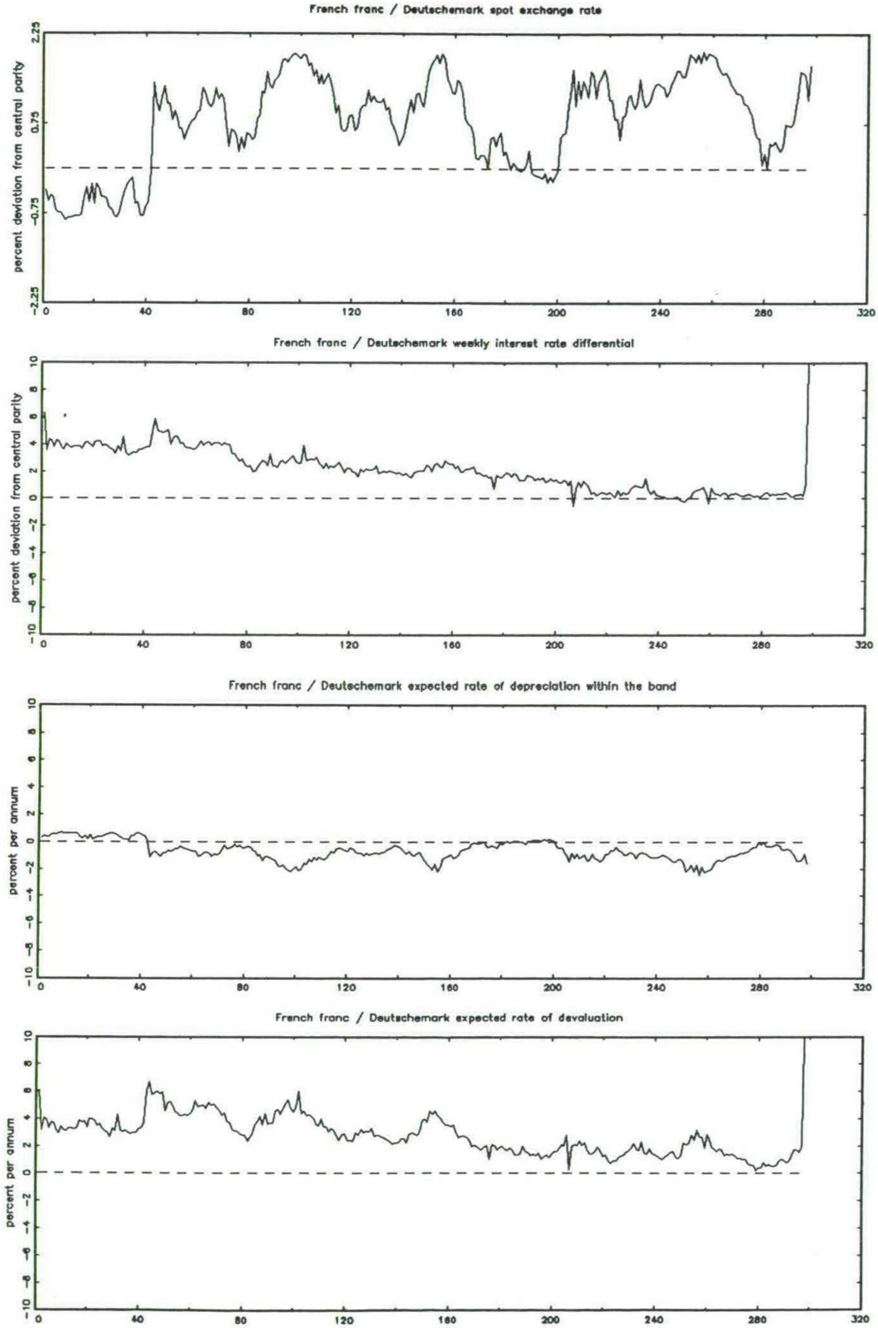


Figure 4.4D Data and expected rate of realignment for French franc / Deutschemark exchange rate.



4.5 Conclusion.

In this chapter we developed a discrete time model for EMS exchange rates. The model has several advantages over continuous time models. First, the distributional assumptions are flexible; the latent fundamental is not restricted to have normally distributed innovations. Second, the model is easily solved numerically. Having solved the model, the likelihood function and moments of the predictive distribution of the exchange rate are immediately available, and standard econometric methods can be used to estimate and test the model.

The empirical results generally support the model. Mean reversion of the fundamental is significant, which suggests that intra-marginal interventions are important in the EMS. The deviation from normality is empirically important as well; a Student-*t* distribution for the innovations in the fundamental is strongly preferred over the normal. We also tested the credibility of the band in the stable EMS period from 1987 to 1992. Although there have been no realignments in that period, the empirical results show that the band has never been fully credible, except for the Deutschemark/Dutch guilder exchange rate.

Appendix. Existence of a unique solution.

In order to prove that (4.6) is a contraction, define the operator T by

$$Tg(f) = f + \lambda\int g(v)dP_{\varepsilon}(v|f)$$

We shall prove that T satisfies Blackwell's sufficient conditions for a contraction (Sargent (1987, p.344)):

(i) Monotonicity: for every $g \geq h$, $Tg \geq Th$.
(ii) Discounting: for every $a > 0$, $T(g+a) \leq Tg + \beta a$, with $0 \leq \beta < 1$.

ad (i)
$$\begin{aligned} Tg(f) &= f + \lambda\int g(v)dP_{\varepsilon}(v|f) \\ &= f + \lambda\int h(v)dP_{\varepsilon}(v|f) + \lambda\int [g(v)-h(v)]dP_{\varepsilon}(v|f) \\ &\geq f + \lambda\int h(v)dP_{\varepsilon}(v|f) = Th(f). \end{aligned}$$

Here we used the fact that the expectation, taken over the distribution function $dP_{\varepsilon}(v|f)$, of the positive function $g(v)-h(v)$ is always positive.

ad (ii)
$$\begin{aligned} T(g(f)+a) &= f + \lambda\int [g(v)+a]dP_{\varepsilon}(v|f) \\ &= f + \lambda\int g(v)dP_{\varepsilon}(v|f) + \lambda a \\ &= Tg(f) + \lambda a, \quad 0 \leq \lambda < 1. \end{aligned}$$

Part II: Stock Markets

Chapter 5

A Comparison of the Cost of Trading French Shares on the Paris Bourse and on SEAQ International

5.1 Introduction.

The growing importance of London as an international stock market where shares from other European countries are traded, constitutes a major change in the structure of Europe's financial markets. In recent years, London's SEAQ International has attracted considerable trading volume from the continental exchanges (see for example Worthington (1991)). This increased competition from London has induced the domestic exchanges to modernise and adapt their trading systems. An example is the move towards fully automated trading systems in Spain and Italy.

It seems natural to suppose that London has attracted large volume because trading costs are lower, particularly for large trade sizes. In this chapter we use a large data set, a simultaneous record of all quotes, limit orders and transactions in both London and Paris, to compare the implicit cost of trading French shares on the Paris Bourse and on SEAQ International. The bid-ask spread is a major component of the total cost of trading, and we will provide a number of measures of the spread on both exchanges. We also briefly discuss published information on explicit costs of trading (such as commissions) in order to gauge the total cost of trading.

In the chapter two different types of estimates of the bid-ask spread are presented. First, the average **quoted** spread is estimated from the Paris limit order book and market makers' quotes in London. Second, the average **realised** spread is estimated from actual transaction prices. Estimates of the quoted and realised spread are presented for different times of day and for different transaction sizes. The dependence of the spread on trade size is also of theoretical interest, because it can be used to assess the validity of market micro-structure theories that predict that the bid-ask spread will be increasing in trade size.

Both the quoted and the realised spread are not directly observable in our data set. On the Paris Bourse part of limit order can be hidden from the public information system, so that the limit order book seems less deep than it actually is. Uncorrected estimates would therefore overestimate the quoted spread in Paris. In London the problem is that there is some misreporting of transaction times, which causes a timing bias in our realised spread estimate. In order to circumvent these problems we also present model-based estimates of the average realised spread using transaction prices only. These estimators can be seen as refinements of Roll's (1984) estimator.

The setup of the chapter is as follows. In section 2 we briefly discuss the major theories that explain the existence and the size of the bid-ask spread. In section 3, we describe the trading systems on the Paris Bourse and on SEAQ International. In section 4 we describe our data. The spread estimates are presented in sections 5, 6 and 7. In section 5 we compute the average quoted spread and in section 6 the average realised spread, both in Paris and in London. In section 7 we take a model-based approach to estimating the realised spread that uses transactions data only. We summarise the main conclusions in section 8.

5.2 Theories of the bid-ask spread.

In the literature on stock market micro-structure there are a number of theories that explain the bid-ask spread. Most theories view the spread as a compensation for the services of a market maker, who takes the other side of all transactions. In the literature, e.g. Stoll (1989), three cost components are distinguished: order processing cost (including dealer oligopoly profit), inventory control cost and adverse selection cost. In this section, these three components will be discussed in more detail.

The order processing cost component reflects the cost of being in the market and handling the transaction. To compensate for these costs, the market maker levies a fee on all transactions by differentiating between buy and sell prices. Much of the empirical literature, such as Madhavan and Smidt (1991) and Glosten and Harris (1987), assumes that this fee is a fixed amount per share. However, it seems more natural to suppose that order processing cost is largely fixed *per transaction*, so that expressed as cost per share it should be declining in trade size.

A second type of cost for the market maker is the cost of inventory management. For example, a purchase of shares will raise the market maker's inventory above a desired level. The market maker runs the risk of price fluctuations on his inventory holdings and if he is risk averse he will demand a compensation for this risk. This intuition is formalised in the model of Ho and Stoll (1981), who show that the inventory control cost is an increasing function of trade size and share price volatility.

The third type of cost for the market maker arises in the presence of asymmetric information between the market maker and his potential counterparties in trading. This theory was first proposed by "Bagehot" (1971) and formalised in the models of Glosten and Milgrom (1985) and Kyle (1985). A trader with superior private information about the underlying value of the shares will try to buy or sell a large number of shares to reap the profits of this knowledge. The market maker, who is obliged to trade at the quoted prices, incurs a loss on transactions with better informed counterparties. To compensate for this loss he will charge a fee on every transaction, so that expected losses on trades with informed traders are compensated with expected profits on transactions with uninformed "noise" traders. Because the informed parties would tend to trade a large quantity in order to maximise the profits from trading on superior information, the adverse selection effect is related to trade size: large transactions are more likely to be initiated by informed traders than small transactions, as in the model of Easley and O'Hara (1987). Therefore, the asymmetric information cost is an increasing function of trade size, and the market maker's quotes for large transactions will be less favourable than the quotes for small sizes.

These theories have been developed for markets with competitive designated market makers. Nevertheless, the theories are frequently applied to exchanges with different trading systems, such as the NYSE. The trading system on the Paris Bourse also differs considerably because there are no designated market makers. But we may regard the issuers of public limit orders as market makers because they provide liquidity to the market. Just like market makers they run the risk that their limit order will be executed against a market order placed by somebody with superior information. The inventory control theory is applicable to the extent that we can regard those who place market orders as demanders of immediacy, while those who place limit orders are making the market by absorbing inventories in return for a price concession. In practice, the distinction between the two groups is not sharp, as any trader can place both types of orders.

Summarising, processing costs cause a decreasing, whereas both asymmetric information and inventory control cause an increasing spread as a function of trade size. The aim of this chapter is to compare the depth of the market in Paris and London and to estimate the form of the dependence of the spread on trade size from quote and transactions data. We shall not attempt to identify the relative importance of the three cost components in this chapter. This decomposition is the focus of Chapter 6, where we concentrate on the dynamic effects of trades on quotes and transaction prices.

5.3 Description of the markets in French equities.

In this section we describe the trading systems on the major exchanges where French equities are traded: the Bourse in Paris and SEAQ International in London. Because the trading systems are so different - Paris is a continuous auction market whereas London is a dealership market - we devote two separate sub-sections to this description.

5.3.1 The trading system on the Paris Bourse.

The Paris Bourse uses a centralised electronic system for displaying and processing orders, the Cotation Assistée en Continu (CAC) system. This system, based on the Toronto Stock Exchange's CATS (Computer Assisted Trading System), was first implemented in Paris in 1986. Since then, trading in nearly all securities has been transferred from the floor of the exchange onto the CAC system. All the most actively traded French equities are traded on a monthly settlement basis in round lots of 5 to 100 shares set by the Société des Bourses Françaises (SBF) to reflect their unit price. The SBF itself acts as a clearing house for buyers and sellers, providing guarantees against counterparty default.

Every morning at 10 a.m. the trading day opens with a batch auction where all eligible orders are filled at a common market clearing price. Nowadays the batch auction is relatively unimportant, accounting for no more than 10 to 15% of trading volume. Its role is to establish an equilibrium price before continuous trading starts. Continuous trading takes place from 10 a.m. to 5 p.m.

In the continuous trading session there are two types of orders possible, limit orders and market orders. Limit orders specify the quantity to be bought or sold, a required price and a date for automatic withdrawal if not executed by then, unless the limit order is good till cancelled ("à révocation"). Limit orders cannot be issued at arbitrary prices because there is a minimum "tick" size of FF 0.1 for stock prices less than FF 500, and FF 1 for higher prices. More than one limit order may be issued at the same price. To these orders, strict time priority for execution applies.

Market orders only specify the quantity to be traded and are executed immediately "au prix du marché", i.e. at the best price available. If the total quantity of the limit orders at this best price does not suffice to fill the whole market order, the remaining part of the market order is transformed into a limit order at the transaction price (for a detailed description of this system see Biais et al. (1992)). Hence, market orders do not automatically walk up the limit order book, and do not always provide immediate execution of the whole order[1].

After the opening, traders linked up to the CAC system will see an onscreen display of the "market by price" as depicted in Figure 5.1.

Figure 5.1 Simplified trading screen of CAC system.

Accor, 24-5-1991, 10:08

Bid			Ask			Transactions		
#	shares	price	price	shares	#	shares	price	time
1	200	763	770	800	3	400	765	10:08
1	500	762	774	100	1	50	765	10:08
1	400	761	775	200	1	50	770	10:06
4	450	760	778	1000	1	50	770	10:02
1	50	754	779	100	1	100	768	10:02

Note: # denotes the number of limit orders involved.

For both the bid side and the ask side of the market, the five best limit order prices are displayed together with the quantity of shares available at that price and the number of individual orders involved. The difference between the best bid and ask price is known as the "fourchette". Traders can scroll down to further pages of the screen to view limit orders available beyond the five best prices. In addition, some information concerning the recent history of trading is given: time, price, quantity and buyer and seller identification codes for the five last transactions, the cumulative quantity and value of all transactions since the opening, and the price change from the previous day's close to the latest transaction.

In practice, the underlying limit order book tends to be somewhat deeper than suggested by the visible display of limit orders. This is because traders who are afraid that they might move the market by displaying a very large order may choose to display only part of their limit order onscreen. The remaining part, known as the "quantité caché" or undisclosed quantity, remains invisible onscreen but may be

[1] A trader who wants to trade a certain quantity immediately can circumvent this mechanism by placing a *limit order* at a very unfavourable price. This limit order will then be executed against existing orders on the other side of the market that show a more favourable price.

called upon to fill incoming orders as the visible limit orders become exhausted. Strict price priority applies also to the hidden orders. Röell (1992) suggests that due to the quantité caché the visible depth of the market is about two thirds of the actual depth when hidden quantities are included.

The member firms of the Bourse (the "Sociétés de Bourse") key orders directly into the CAC system via a local terminal. All market participants can contribute to liquidity by putting limit orders on display. In particular, the Sociétés de Bourse may act in dual capacity: as agency brokers, acting on behalf of clients, and as principals, trading on own account. Their capital adequacy is regulated and monitored by the Bourse.

There is some scope for negotiated deals if the limit order book is insufficiently deep. A financial intermediary can negotiate a deal directly with a client at a price lying within the current fourchette, provided that the deal is immediately reported to the CAC system as a "cross order". For trades at prices outside the fourchette, the member firm acting as a principal is obliged to fill all central market limit orders displaying a better price than the negotiated price within five minutes.

5.3.2 The trading system on SEAQ International.

SEAQ International is the price collection and display system for foreign equity securities operated by London's Stock Exchange. For each foreign equity included in SEAQ International, the system provides an electronic display of bid and ask prices quoted by the market makers registered for that equity.

The French equities in our sample are designated as firm quote securities, which means that during the relevant mandatory quote period (9:30 to 16:00 London time, i.e. 10:30 to 17:00 Paris time in our sample) the registered market makers are obliged to display firm bid and ask prices for no less than the "minimum marketable quantity", also referred to as the Normal Market Size (NMS), a dealing size set by the exchange's Council at about the median transaction size. Market makers are obliged to buy and sell up to that quantity at no worse than their quoted prices. In addition, when a market maker displays a larger quantity of shares than the minimum marketable quantity, his prices must be firm for that quantity. Outside the mandatory quote period, market makers may continue to display prices and quantities under the same rules regarding firmness of prices.

SEAQ International market makers are not allowed to display prices on competing display systems which are better than those displayed on SEAQ International. Market making in French shares is fairly competitive, see Röell

(1992): during our sample period, most French equities were covered by at least ten market makers, and usually many more.

5.4 Data description.

In this section we describe the data provided to us by the Paris and London exchanges. We have quote and transaction data from both exchanges for the same period in the summer of 1991.

5.4.1 The Paris Bourse.

The Paris data set is a transcription of all changes in the trading screen information for all shares on the CAC system for 44 trading days in the summer of 1991, starting May 25 and ending July 25. We have available a complete record of the total limit order quantity at the five best prices on both the bid side and the ask side of the market and all transactions. This enables us to reconstruct at every point in time the visible limit order book for each security in the sample, up to the cumulative volume of the observed best limit orders. However, we do not observe the "quantité caché", so the actual limit order book might be deeper than the observed quantities suggest.

Concerning transactions, there is an indicator showing whether the transaction is a "cross" negotiated outside the CAC system. We also have available broker identification codes of the buying and selling parties, which allow us to identify series of small transactions that were initiated by the same person as part of one large transaction. The transaction price per share for such transactions is defined as the quantity weighted average of the price of the small transactions that together make up the larger one.

In this chapter we shall concentrate on ten major French stocks, listed and described in Table 5.1A. This table also shows some descriptive statistics of our sample of transactions on the Paris Bourse. The table shows that (excluding cross transactions) the median transaction value is between FF 50,000 and FF 150,000 (£5,000-£15,000 at the time). The distribution of transaction size is very skewed: the mean is about twice the median, indicating that a few large transactions account for a large share of total turnover. The table also shows descriptive statistics for the "cross" transactions that are negotiated off-exchange. The crosses are relatively large: their median value is about 2 to 5 times as large as the median value of regular transactions, and the mean value is up to 10 times the mean value of regular transactions. Although there are relatively few crosses (between 2 and 5 % of the total number of transactions) they account for a large share of total trading volume.

Table 5.1A Descriptive statistics of Paris transactions data.

			CAC			crosses		
Firm	full name	average price	median value	mean value	nobs	median value	mean value	nobs
AC	Accor	771	114	197	5255	384	2531	148
AQ	Elf-Aquitaine	358	179	303	9855	183	1607	598
BN	BSN	889	62	182	10728	266	1039	378
CA	Carrefour	1919	90	164	9943	366	1268	307
CS	Axa-Midi	989	62	120	6482	89	2266	221
EX	Generale des Eaux	2518	129	247	9585	366	2070	475
OR	l'Oreal	584	64	145	6813	116	694	271
RI	Pernod-Ricard	1162	84	131	3626	327	838	123
SE	Schneider	685	68	134	4329	388	876	183
UAP	Un. Ass. de Paris	538	134	222	5206	54	728	402

Notes: price is average transaction price in FF;
value of transactions in FF1000; nobs is number of observations.

5.4.2 SEAQ International.

The data from the London exchange cover May to July 1991. Table 5.1B shows some statistics for the ten stocks under consideration. There are fewer transactions in London than in Paris, but the median size of the transactions is much larger. The NMS is generally valued at about FF 1 million (£100,000), a rather large transaction by Paris standards. The average value of transactions in London is about 10 times the average value of regular transactions in Paris, and still somewhat larger than the mean value of crosses in Paris.

Table 5.1B Descriptive statistics of London transactions data.

Firm	median value	mean value	nobs	NMS
AC	1094	2049	393	2000
AQ	1473	2966	1168	5000
BN	862	1487	853	2500
CA	950	2293	771	500
CS	758	1858	291	1000
EX	1106	2545	905	500
OR	732	1691	449	2500
RI	630	1479	210	1000
SE	1100	1970	204	2000
UAP	1532	2460	518	2000

Notes: value expressed in FF1000; nobs is number of transactions is sample
NMS is Normal Market Size in number of shares

Percentiles of the distribution of trade size in London and Paris are given in Table 5.2. The results indicate that there are many very large transactions in London compared with Paris. For example, the 99^{th} percentile in Paris is in the order of magnitude of one NMS, whereas the 90^{th} percentile in London already is about 5 times NMS.

Table 5.2 Percentiles of transaction size distribution.

	Paris					London				
%	90	95	99	99.5	99.9	90	95	99	99.5	99.9
AC	0.25	0.50	1.0	1.4	2.5	2.5	4.4	12.4	15.0	20.0
AQ	0.40	0.60	2.0	3.0	16.2	3.0	4.6	15.7	22.0	50.0
BN	0.21	0.36	0.8	1.2	2.4	1.6	2.4	5.0	6.0	12.0
CA	0.40	0.60	1.6	2.1	5.0	5.0	8.0	30.0	40.0	51.0
CS	0.29	0.48	1.1	3.0	5.5	4.0	6.0	13.3	21.6	32.8
EX	0.46	0.80	2.0	4.0	12.0	4.2	7.5	18.0	26.7	72.8
OR	0.22	0.39	1.0	1.5	2.8	2.2	3.4	9.8	18.8	20.2
RI	0.25	0.43	1.2	1.8	3.0	3.3	5.1	8.8	8.8	9.1
SE	0.25	0.48	1.0	1.5	2.5	3.8	5.2	9.7	10.0	10.7
UAP	0.20	0.75	2.5	4.6	5.5	5.0	7.5	12.5	15.5	21.2

Notes: percentiles expressed in NMS; crosses included in Paris sample.

5.4.3 Distribution of transactions by time of day.

Figure 5.2 depicts the distribution of trades by time of day in Accor shares in Paris. In this figure the trading day is split up into seven hours, ranging from 10:00am to 17:00pm (the period of continuous trading) and the number of transactions in each interval for all days in the sample is counted. There is a clear lunch break effect between 1pm and 2pm. More interestingly, most trading takes place in the hour after opening and the hour before closing. The graphs for the other series show very similar patterns. Hence, our data match the U-shaped trading pattern found by McInish and Wood (1990) for the Toronto Stock Exchange, by Niemeyer and Sandas (1992) for the Stockholm exchange and by Kleidon and Werner (1993) for the S&P 100 firms on the NYSE. In contrast, Schmidt and Iversen (1991) found an inverted U-shaped trading pattern with the trading sessions of the German MATIS.

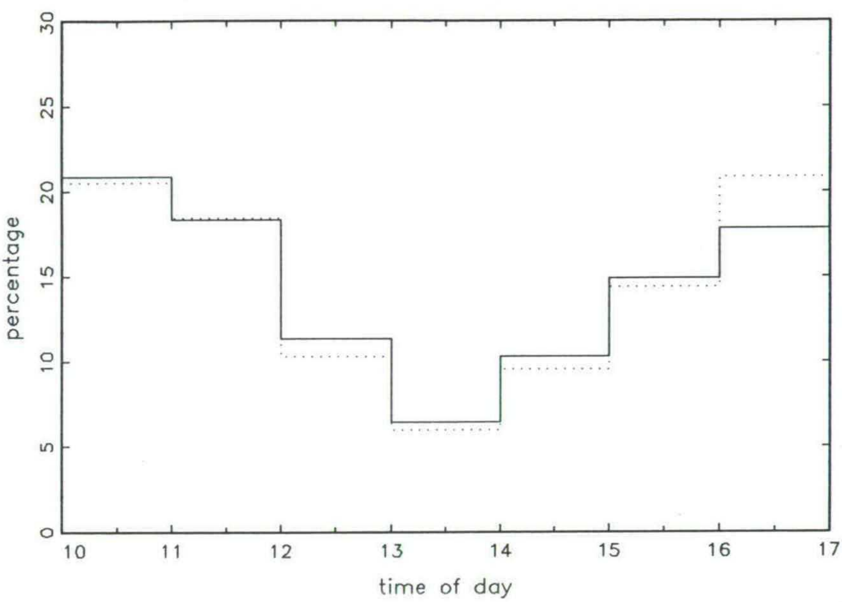


Figure 5.2: Trading volume by time of day Accor

5.5 The quoted spread for French equities.

In this section we provide an analysis of the cost of immediacy on the Paris Bourse and SEAQ International. The cost of an urgent transaction is determined by the available orders in the limit order book in Paris and by the market maker quotes in London. An important determinant of the cost of immediacy therefore is the quoted spread. For Paris, the average quoted spread is determined as the average difference between bid and ask prices in the limit order book for a certain size. In London, the quoted spread is the difference between the best bid and ask quotes of the market makers. Although prices are negotiable in London, one cannot always count on "within-the-touch" prices for an immediate transaction.

In order to compute the quoted spread in Paris it is necessary to construct the limit order book. We observe all new limit orders, as well as all transactions that fill limit orders and orders that are withdrawn, so that we can recursively build up the order book over the day. There are two problems in constructing the order book, however. First, there is the unobserved "quantité caché", which makes the book deeper than observed. Second, we observe only the limit orders at the five best prices, so that we do not have prices for larger order sizes. In constructing the book we impute the fifth best limit order price for all sizes beyond the range for which the bid and ask price are observed [2]. Clearly this biases the average quoted spread downwards. On the other hand, ignoring the quantité caché biases the average upwards. The net effect of these data imperfections on the estimates of the quoted spread is indeterminate.

A first way to measure the average quoted spread is by a simple **calendar time average** of the observed spreads between bid and ask prices:

$$(5.1) \qquad S_C^m(z) = \Sigma_{i=1}^N (t_{i+1}-t_i)\Big(A[t_i,z]-B[t_i,z]\Big) \;/\; \Sigma_{i=1}^N (t_{i+1}-t_i),$$

where t_i is the calendar time index (in seconds) of the i^{th} change in the limit order book, $A[t_i,z]$ denotes the marginal ask price at time t_i for an order of size z, $B[t_i,z]$ denotes the corresponding bid price, and N is the number of quote changes in the sample period. Thus, each quote is weighted in proportion to the calendar time for which it appears on the trading screens (that is, the time until the next quote change). This measure computes the quoted price for the marginal unit only. The *average* spread for a hypothetical transaction of size z is obtained by averaging the marginal spread over all smaller sizes:

$$(5.2) \qquad S_C(z) = \Sigma_{i=1}^N (t_{i+1}-t_i)\Sigma_{y=1}^z z^{-1}\Big(A[t_i,y]-B[t_i,y]\Big) \;/\; \Sigma_{i=1}^N (t_{i+1}-t_i).$$

For London, the average quoted spread or "market touch" can be derived directly from the market maker's quotes. The London quotes apply to all sizes smaller than or equal to NMS.

Table 5.3 hows the average quoted spread in London and Paris for sizes up to NMS. The averages are taken over the period of continuous trading in Paris and the mandatory quote period in London. All quotes outside normal trading hours are ignored. There is no point in going beyond NMS because the data on the limit order

---

2 An alternative procedure is to exclude those observations for which we do not observe the quoted bid and ask price up to the required size. That procedure introduces a selection bias in the spread measure because the five best limit orders add up to a large size only when the market is deep. Hence, that procedure underestimates the spread. Comparison of this alternative procedure with the procedure described in the main text showed that the selection bias is more serious than the bias caused by imputing the fifth best price for unobserved limit orders. See also Anderson and Tychon (1993) who report large selection biases for Belgian stocks.

book in Paris for larger sizes are very sparse and therefore spread estimates for large sizes are unreliable.

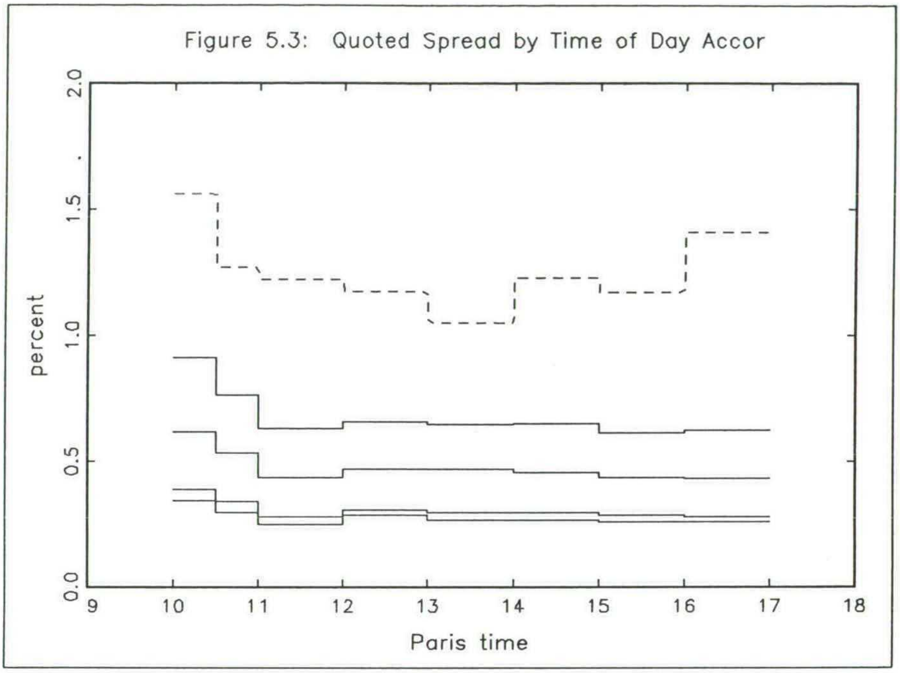
Table 5.3 Calendar time average quoted spread.

	Paris				London
size	0.0	0.1	0.5	1.0	≤1.0
AC	0.287	0.277	0.444	0.646	1.268
AQ	0.187	0.206	0.300	0.409	0.977
BN	0.176	0.218	0.359	0.519	0.852
CA	0.216	0.245	0.372	0.504	1.346
CS	0.363	0.427	0.655	0.898	2.309
EX	0.135	0.147	0.210	0.285	1.004
OR	0.308	0.383	0.672	0.980	1.646
RI	0.367	0.413	0.650	0.912	2.106
SE	0.370	0.452	0.838	1.200	2.061
UAP	0.422	0.467	0.685	0.972	1.712

Notes: percentage quoted spread by $S_C(z)$ definition.

The table reveals that the calendar time average of the spread, $S_C(z)$, in Paris is very small for small sizes, between 0.15-0.4%, and rises for larger sizes. For example, at NMS the average quoted spread in Paris is between 0.3 and 1.2%. The quoted spread in London (the "touch") is considerably bigger than the quoted spread in Paris for all sizes below NMS. The quoted spread for NMS in Paris is only half of the quoted spread in London. Röell (1992), who had available a number of snapshots of the complete limit order book (including the quantité caché) concludes that the Paris quotes are narrower than the London quotes for order sizes of up to 2 times the NMS, but that for larger sizes the limit order book runs out quickly.

Figure 5.3 shows a breakup by time of day of the average quoted spread for Accor shares. For Paris, the average fourchette (the difference between best bid and ask prices at the smallest lot size) and the quoted spread for 0.1, 0.5 and 1 times NMS are shown. For London, the average "touch" is graphed. The graph shows hourly time intervals, except for the early morning hour which is split in two because the London mandatory quote period starts only at 10.30 a.m. Paris time. It is clear that the Paris market is very tight, the fourchette is only about 0.25% for the series plotted and doesn't change much over the day, although it is slightly higher in the first half hour, which falls outside the mandatory quote period in London. The quoted spread in London, graphed as the dotted line in Figure 5.3, is much larger than the fourchette and also larger than the average quoted spread at NMS in Paris for all times of the day.



An obvious drawback of the above spread measure is that it is a calendar time average, hence periods in which there is hardly any trading are given the same weight as periods of equal length in which trading is heavy. A second estimator conditions on the actually observed trade pattern by taking a **transaction time average** of the difference between bid and ask prices:

$$(5.3) \qquad S_T(z) = \frac{1}{N} \Sigma_{i=1}^{N} \Sigma_{y=1}^{Z} \Big(A[t_i,y]-B[t_i,y]\Big)/z,$$

where t_i denotes the calendar time index of transaction i.

There are no large differences between the calendar time average, $S_C(z)$, and the transaction time average, $S_T(z)$, reported in Table 5.4. For Paris, the transaction time average quoted spread is slightly smaller than the calendar time average at small sizes, and about the same at NMS. For London, the calendar time and transaction time average quoted spreads are also very similar. The timing of transactions therefore does not seem to be very sensitive to variations in the spread. This is also evident from a comparison of Figures 5.2 and 5.3: trading volume does not seem to be concentrated at times of day when the fourchette is particularly narrow. Trading volume is U-shaped over the day, while the fourchette does not display an inverted U-shape[3]. Therefore, the transaction time average of the quoted spread is rather similar to the calendar time average.

Table 5.4 Transaction time average of the percentage quoted spread.

	Paris				London	truncation		
size	0.0	0.1	0.5	1.0	≤1.0	0.1	0.5	1.0
AC	0.228	0.258	0.427	0.629	1.315	0	0	18
AQ	0.174	0.197	0.305	0.426	0.954	0	2	45
BN	0.174	0.219	0.386	0.531	0.852	0	2	22
CA	0.195	0.223	0.349	0.478	1.228	0	4	33
CS	0.336	0.405	0.655	0.913	2.208	0	6	50
EX	0.130	0.143	0.211	0.290	1.006	0	1	13
OR	0.293	0.369	0.673	0.987	1.624	0	7	50
RI	0.339	0.394	0.649	0.921	2.159	0	5	56
SE	0.334	0.414	0.782	1.150	2.025	0	11	66
UAP	0.390	0.436	0.665	0.959	1.685	0	0	17

Notes: percentage quoted spread by $S_C(z)$ definition;

"truncation" gives the percentage of observations for which the size the observed limit orders didn't add up to 0.1, 0.5 or 1 NMS, respectively, on either the bid or the ask side (or both).

A further refinement of the quoted spread measure is obtained if we condition not only on the pattern of trades over the day, but also on the size of transactions. The results of Biais et al. (1992) suggest that indeed large transactions tend to take place at times when it is relatively cheap to trade large quantities. This is formalised in the third estimator, which averages the quoted spread over times that transactions in a particular size class occurred:

$$(5.4) \qquad S_Q(\underline{z},\bar{z}) = \sum_{i=1}^{N} I(\underline{z}<z_i\leq\bar{z})\sum_{y=1}^{z_i}\Big(A[t_i,y]-B[t_i,y]\Big)/z_i \;/\; \sum_{i=1}^{N} I(\underline{z}<z_i\leq\bar{z}),$$

where I(.) is an indicator function that takes the value one if the trades size exceeds the lower bound $\underline{z}$ and is smaller than or equal to the upper bound, $\bar{z}$, and takes the value zero otherwise. Table 5.5 reports the quoted spread S_Q for several size classes.

[3] Schmidt and Iversen (1991) did find a clear U shaped spread pattern that was just the opposite of the inverted U shaped trading pattern.

Table 5.5 Transaction time average of percentage quoted spread by size class.

A. Paris.

	≤0.1	0.1-0.5	0.5-1	>1.0	all
AC	0.237	0.271	0.472	0.523	0.252
AQ	0.179	0.218	0.324	0.363	0.201
BN	0.182	0.228	0.422	0.555	0.197
CA	0.209	0.234	0.336	0.375	0.225
CS	0.356	0.434	0.697	0.720	0.389
EX	0.134	0.154	0.225	0.263	0.148
OR	0.309	0.396	0.706	0.782	0.342
RI	0.359	0.404	0.662	0.778	0.378
SE	0.356	0.449	0.831	0.993	0.386
UAP	0.421	0.465	0.716	0.925	0.453

B. London.

	≤0.1	0.1-0.5	0.5-1	1-2	2-5	>5	all
AC	1.325	1.346	1.336	1.241	1.301	1.345	1.315
AQ	0.995	0.953	0.961	0.960	0.897	0.972	0.954
BN	0.880	0.852	0.845	0.830	0.812	0.849	0.852
CA	1.260	1.245	1.276	1.219	1.160	1.183	1.228
CS	2.418	2.206	2.285	2.132	2.054	2.166	2.208
EX	0.957	0.974	1.061	1.008	0.996	1.075	1.006
OR	1.505	1.652	1.625	1.656	1.680	1.716	1.624
RI	2.208	2.130	2.253	2.201	2.095	1.909	2.159
SE	2.003	2.072	1.954	1.972	2.158	1.922	2.025
UAP	1.841	1.655	1.740	1.681	1.706	1.488	1.685

Notes: percentage quotes spread by $S_Q(z,z)$ definition.

For Paris, the value of S_Q is usually slightly smaller than the value of S_T, indicating that indeed the size of transactions is related to the quoted spread. Like S_C and S_T, S_Q is increasing in trade size, nearly doubling from the smallest to the largest size class. For London, only the "touch" was averaged by transaction size class, and these show no clear pattern. In London, therefore, trade size does not seems to depend on the "touch".

In summary, the results of this section show that the quoted spread in Paris is much smaller than the quoted spread in London for small transaction sizes below NMS. However, for larger transactions the quoted spread in Paris rises quickly as the limit order book runs out. Some care has to be taken with these results because the estimates of the quoted spread in Paris ignore the hidden quantities and are marred by the problem that we only have data on the five best limit orders.

5.6 Realised spread.

In this section we compute spread estimates that are based on transaction prices rather than on quoted prices and will therefore be referred to as measures of the realised spread. The limit order and quote data are used to construct a measure of the mid-price of the stock only. The estimator of the realised spread that we propose is twice the average absolute difference between the quoted mid-price and the transaction price:

$$(5.5) \qquad S_R(\underline{z},\bar{z}) = 2\Sigma_{i=1}^{N} I(\underline{z} < z_i \leq \bar{z}) \cdot |p[i]-m[i]| \ / \ \Sigma_{i=1}^{N} I(\underline{z} < z_i \leq \bar{z}),$$

where as before I(.) is the indicator function, p[i] is the actual transaction price (average price paid per share) and m[i] is the mid-price at the time of the i^{th} transaction, defined as the average of the best bid and ask quote (or best buy and sell limit orders) for the smallest possible order size.

For Paris, there are at least two important differences between the realised spread measure and the quoted spread measures of the previous section. The first is that the limit order book data are required only to construct the mid-price. This means that the realised spread estimate in Paris is not affected by the quantité caché and the availability of only the five best limit order prices. The second important difference is that the implicit assumption that the market is equally deep on both sides is dropped. One would expect that large trades are more likely to take place on the deeper side of the market. If so, the realised spread measure should be lower than the quoted spread measure for larger trade sizes. In London transactions are routinely priced within the touch, and therefore the quoted spread will be an overestimate of the realised cost of trading.

The realised spread is the measure that is relevant for a patient trader, who can wait for the best moment to trade, but it is probably not a good indicator of the cost of immediacy. First of all, an impatient trader cannot choose the deeper side of the market. Moreover, in London transaction prices for larger deals are generally negotiated within the touch, and the same is true for Paris cross transactions. This means that the realised spread, which measures the average cost of actual transactions, understates the cost of a hypothetical urgent transaction, where the trader cannot rely on negotiating within the quote prices.

The estimates S_R of the average realised spread are reported in Table 5.6. In calculating the estimates we excluded all transactions outside the continuous trading (Paris) or the mandatory quote period (London) because outside normal trading hours the mid-quote is not a reliable proxy for the market consensus valuation of the stock.

Table 5.6 Average percentage realised spread S_R

A. Paris.

	≤0.1	0.1-0.5	0.5-1	>1.0	all
AC	0.245	0.236	0.251	0.230	0.242
AQ	0.193	0.202	0.188	0.160	0.196
BN	0.187	0.188	0.201	0.181	0.187
CA	0.227	0.212	0.221	0.200	0.221
CS	0.372	0.378	0.429	0.327	0.375
EX	0.151	0.154	0.158	0.145	0.153
OR	0.325	0.315	0.305	0.171	0.322
RI	0.368	0.352	0.384	0.401	0.364
SE	0.362	0.359	0.311	0.178	0.361
UAP	0.458	0.416	0.434	0.381	0.438

Note: Average based on all transactions in continuous trading period (10-17).

B. London.

	≤0.1	0.1-0.5	0.5-1	1-2	2-5	>5	all
AC	1.357	1.164	1.055	1.286	1.382	0.688	1.191
AQ	1.759	1.237	1.337	1.028	1.298	2.046	1.300
BN	1.495	0.993	1.072	1.296	1.364	5.273	1.202
CA	1.720	1.521	1.137	1.158	1.226	1.390	1.323
CS	3.032	1.856	1.330	1.177	1.668	1.987	1.728
EX	1.452	1.105	1.067	1.035	1.128	1.478	1.148
OR	2.098	1.166	1.456	1.422	1.870	1.739	1.447
RI	2.146	1.192	0.980	1.783	1.410	1.101	1.409
SE	1.374	1.695	1.574	1.744	1.025	1.008	1.503
UAP	1.664	1.242	1.150	1.429	1.371	1.295	1.313

Notes: Transactions only in mandatory quote period (9.30-16).

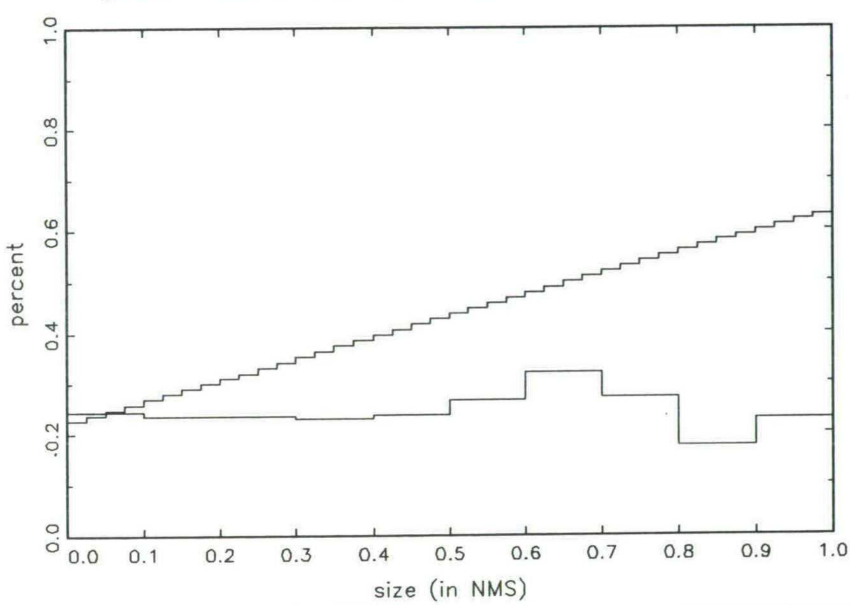
C. London, bias corrected.

	≤0.1	0.1-0.5	0.5-1	1-2	2-5	>5	all
AC	1.339	1.169	1.169	1.286	1.382	0.688	1.191
AQ	0.963	1.360	0.963	1.028	1.298	2.046	1.300
BN	0.857	1.134	0.857	1.296	1.364	5.273	1.202
CA	1.260	1.404	1.260	1.158	1.226	1.390	1.323
CS	2.270	1.859	1.859	1.177	1.668	1.987	1.728
EX	0.998	1.153	0.998	1.035	1.128	1.478	1.148
OR	1.607	1.394	1.394	1.422	1.870	1.739	1.447
RI	2.182	1.354	1.345	1.783	1.410	1.101	1.409
SE	2.016	1.601	1.601	1.744	1.025	1.008	1.503
UAP	1.709	1.232	1.232	1.429	1.371	1.295	1.313

Note: Transactions only in mandatory quote period (9.30-16)
Bias correction described in Appendix.

Table 5.6, panel A shows the average realised spread in Paris. All transactions within the continuous trading period were used, including "crosses". The table clearly shows that the realised spread in Paris does not increase with trade size. In contrast, in the previous section we have seen that the quoted spread increases with size. The dependence of the quoted and realised spread in Paris on trade size is illustrated in Figure 5.4, where a quoted spread estimate (S_T) and the realised spread estimate (S_R) for the Accor series are graphed.

Figure 5.4: Quoted and Realised Spread Accor in Paris by size



Estimates of the average realised spread in London are reported in Table 5.6, panel B. The most striking result here is that the realised spread in London seems to be *declining* in trade size. This effect was also observed by Breedon (1992), Tonks and Snell (1992) and Röell (1992). A comparison of Table 5.4 and Table 5.6 shows that in London the average realised spread for transactions smaller than NMS is generally larger than the quoted spread. This seems impossible: the rules of SEAQ International oblige market makers to stand firm at the best quoted price for transactions smaller than NMS. A likely explanation for this anomaly is a *timing bias*. Unlike in Paris, the reported time of transactions in London can be inaccurate and the market maker quotes are not updated very frequently so that they may well be stale. In the Appendix we show that this biases our realised spread measure upwards, because the market mid-price may have moved between the actual transaction time and the reported time. In Table 5.6, panel C we report bias-adjusted realised spreads for London.

The observation that realised spreads do not increase or even decline with trade size is important because it is not in line with the inventory control and adverse selection models of the spread discussed in section 2, or with the assumption that order processing cost is fixed per share. Constant processing costs per transaction, and therefore declining per share, could be an explanation for the empirical result that the cost *per share* is smaller for large trade sizes. We return to this issue in section 5.7 where we estimate a parametric model for the dependence of the realised spread on trade size.

Comparing the realised spread in London with the realised spread in Paris, it appears that the London realised spread is considerably higher than the Paris one, so Paris seems to be cheaper. There are a number of caveats. First, there are few transactions larger than NMS in Paris (indeed most of those are cross transactions) while in London roughly half the transactions exceed NMS. Second, the full cost of trading also includes taxes and other explicit transaction costs. We return to this point in the concluding section.

5.7 Model-based estimates of the realised bid-ask spread

The spread measures of the previous section relied on data from the limit order book or quotes to construct an estimate of the unobserved consensus value of the stock. The estimators proposed in this section do not require such a proxy, and are therefore less sensitive to the problems encountered in section 5.6. In particular, timing bias is not a problem. The price paid for this improvement is the need to make some parametric assumptions about the process that generates prices. We build some simple models to estimate the realised spread and to estimate the dependence of the spread on trade size.

The simplest model that we consider is based on the work of Stoll (1989) and George, Kaul and Nimalendran (1991). In their models it is assumed that the transaction price, p_t is equal to the (unobserved) mid-price prior to the trade, y_t, plus or minus one-half times the total spread, S. We allow for an error term, u_t, in the price equation, that picks up various effects on the transactions price that are not captured by the mid-price and the transactions type, such as price discreteness and trade size. Thus, the price equation is

(5.6) $$p_t = y_t + (S/2)Q_t + u_t,$$

where Q_t indicates whether the transaction is initiated by the buyer (+1) or the seller (-1). The variable Q_t will henceforth be referred to as the "sign" of the trade. The mid-price y_t in this equation is not observed, so we cannot estimate this model directly. In order to obtain an equation in observables only, we first difference (5.6) and make assumptions on the dynamics of the mid-price.

If there is asymmetric information between market makers and other traders, the market maker will revise his mid-price after a trade has occurred. Moreover, for inventory control reasons he will also change his quotes because the trade changes his inventory. Let $(1-\pi)$ be the fraction of the spread attributable to asymmetric information and inventory control, and π the fraction attributable to processing cost, then the revised mid-price immediately after the transaction is

(5.7) $$m_t = y_t + (1-\pi)(S/2)Q_t.$$

Finally, between two trades public information on the stock's value may come in so that the new mid-price prior to the subsequent trade is a revision of m_t

(5.8) $$y_t = \beta_0 + m_{t-1} + e_t,$$

where where β_0 is the average and e_t is the unexpected mid-price return resulting from public information between the two transactions. Under these assumptions the transaction price returns can be expressed as

(5.9) $$\Delta p_t = \beta_0 + (S/2)Q_t - \pi(S/2)Q_{t-1} + e_t + \Delta u_t.$$

This is an equation in observables (Δp_t and Q_t) and random error terms only. It is a valid regression model under the additional assumption that Q_t is exogenous, so that Q_t and (e_t,u_t) are uncorrelated at all lags[4]. Under this assumption, π and S can be estimated consistently by least squares, where the coefficient of Q_t is half the realised bid-ask spread.

---

[4] If the pricing error u_t is due to rounding, Q_t and u_t might be correlated. In that case, an instrumental variables technique could be used to estimate (5.9). We ignore this point in the estimation.

If we furthermore assume that (e_t, u_t) is a joint white noise process, the regression has a first order moving average error structure. Moreover, the innovations in the true price are probably heteroskedastic, as suggested by the results of Hausman, Lo and MacKinlay (1992). One of the reasons for the heteroskedasticity is the difference in the calendar time span between transactions. However, there may be other factors that cause a time-varying conditional variance. Instead of specifying the form of heteroskedasticity, we estimate by OLS, which under the stated assumptions gives consistent point estimates, and compute heteroskedasticity and autocorrelation consistent (HAC) standard errors using the method proposed by Newey and West (1987).

In the literature several spread estimators have been developed for cases where no data on sign or size of the transactions are available. Roll (1984) proposes an estimator of the spread based on the first order autocovariance of the returns, $\gamma_{\Delta p}=E(\Delta p_t \Delta p_{t-1})$. In the simple model (5.9), Roll's estimator is consistent only under some very restrictive assumptions: no serial correlation in expected returns; no error term in the price equation ($\sigma_u^2=0$); no serial correlation in the transaction type ($E[Q_t Q_{t-1}]=0$); and no asymmetric information or inventory control effects ($\pi=1$). Under these assumptions, the first order autocovariance of the returns is equal to $-(S/2)^2$, and Roll's estimator of the spread is given by

$$\hat{S} = 2\sqrt{-\gamma_{\Delta p}}. \tag{5.10}$$

Roll's estimator is biased downward if there is positive serial correlation in the transaction sign Q_t (i.e. if transactions at the bid tend to be followed by further transaction at the bid and similarly for the ask). Choi, Salandro and Shastri (1988) adjust to Roll's estimator for serial correlation in Q_t, retaining the assumptions that there are no pricing errors ($\sigma_u^2=0$), no serial correlation in mid-price returns and no asymmetric information or inventory control effects ($\pi=1$). Choi et al. (1988) assume also that Q_t follows a first order Markov process. Under these assumptions, the first order autocovariance of the returns is

$$\gamma_{\Delta p} = -(S/2)^2(1-\gamma)^2, \tag{5.11}$$

where γ_Q is the first order autocovariance of the transaction sign. From (5.11), the CSS estimator follows directly

$$\hat{S} = 2\sqrt{-\gamma_{\Delta p}}/(1-\gamma), \tag{5.12}$$

This estimator takes the form of a simple correction by a factor $1/(1-\gamma)$ of Roll's estimator.

The details of the estimation procedures are as follows. In line with the previous sections, we exclude all transactions outside the mandatory quote period (London) and the period of continuous trading (Paris). We include all other transactions, including the crosses in Paris. We take logarithms of the transaction prices and multiply those by 100 to obtain estimates of the percentage spread. The estimation equation is thus specified in returns, but only within-day returns were used because overnight returns are unlikely to follow the same process as intra-day returns, see Harris (1986). The classification of the trade as buyer initiated or seller initiated is done by comparing the transaction price with the mid-price. If the transaction price exceeds the mid-price, the trade is classified as buyer initiated ($Q_t=1$), and if the transaction price is lower than the mid-price the trade is classified as seller initiated ($Q_t=-1$). If the transaction price is exactly at the mid-price, the trade is not classified and the value 0 is assigned to Q_t. This procedure is exact for the Paris transactions that were executed through the CAC system, but for the crosses and the London data there might be some incorrect classifications due to reporting lags.

The model-based estimates of the realised spread in London and Paris are given in Table 5.7. Like our previous results in Section 5.6, the model based estimates suggest that the realised spread in London substantially exceeds the realised spread in Paris. Comparing the average realised spread S_R in Table 5.5 with the regression-based estimated spread, the latter is smaller for all stocks, suggesting that the average of best bid and ask quotes is not a good approximation of the unobserved true mid-price. This discrepancy is particularly striking in the case of the London data. There, because the market maker quotes are updated relatively infrequently, our data on the quoted mid-price may be a particularly inappropriate basis for computing realised spreads.

Table 5.7 Model based estimates of realised spread.

model: $\Delta p_t = \beta_0 + \beta_1 Q_t + \beta_2 Q_{t-1} + \varepsilon_t$

	Paris			London		
	Roll	CSS	$2\beta_1$	Roll	CSS	$2\beta_1$
AC	0.178	0.259	0.214 (47.855)	1.075	1.802	0.890 (10.214)
AQ	0.143	0.196	0.167 (65.196)	1.136	2.040	1.290 (13.740)
BN	0.147	0.182	0.169 (86.701)	0.679	1.354	0.781 (12.666)
CA	0.154	0.241	0.179 (56.733)	1.003	1.991	0.809 (11.010)
CS	0.274	0.359	0.330 (48.947)	2.152	3.997	1.131 (6.961)
EX	0.109	0.157	0.123 (58.959)	0.954	1.717	0.771 (12.077)
OR	0.246	0.328	0.285 (59.805)	0.849	1.401	0.992 (11.418)
RI	0.248	0.336	0.305 (35.139)	0.748	1.284	0.819 (6.765)
SE	0.253	0.371	0.316 (41.965)	2.071	4.186	1.901 (4.396)
UAP	0.349	0.521	0.404 (49.420)	0.902	1.444	0.842 (11.575)

Heteroskedasticity and autocorrelation consistent t-ratios in parentheses;
Estimates based on all transactions within regular trading hours;
Spread expressed as a percentage of the transaction price.

In addition to the effect of the *sign* of the trade (buyer or seller initiated) the *size* of the trade may also be an important determinant of the price. The microstructure theories discussed in section 5.2 predict that due to asymmetric information and inventory control the spread will be an increasing function of trade size. To estimate the effect of size we extend model (5.6) in the spirit of Glosten and Harris (1988) and Madhavan and Smidt (1991). The price equation is extended with a linear term in the size of the transaction. In section 5.5 we found some evidence for a fixed processing cost per transaction that would generate a decreasing processing cost for large trade sizes. This effect is captured by adding the inverse of trade size to the price equation. Together, we add two additional variables to (5.6) and obtain

$$(5.6') \qquad p_t = y_t + (S/2)Q_t + \alpha z_t + \gamma z_t^{-1} + u_t,$$

where z_t is the signed trade size. First differencing (5.6′) we obtain the equivalent of regression equation (5.9) but now including current and lagged trade size and the inverse of size as regressors[5]:

$$(5.9') \qquad \Delta p_t = \beta_0 + (S/2)Q_t + \alpha z_t + \gamma z_t^{-1} + \text{lagged terms} + e_t + \Delta u_t.$$

In order to reduce the influence of very large transactions (outliers) on the estimates, we "censor" large trade sizes. For Paris, we pick the threshold at 2 NMS, which is about the 99.5% quantile[6]. In London many more trades would be censored at 2 NMS, between 10 and 25 percent. Therefore, we use a threshold of 5 NMS for the London data, which corresponds to the 95% quantile.

The estimation results are reported in Table 5.8. For Paris, the coefficients of the size and the inverted size are small but significant for most cases. The Wald test of joint significance of the size and inverted size parameters is larger than its 5% critical value (5.99) for all series except one. On the other hand, for London less evidence for a trade size effect on the realised spread is found. The size effect is jointly significant only for BSN (BN) and Axa-Midi (CS). Partly this may reflect the smaller sample size of the London series.

Hasbrouck (1991a) uses a more extensive model to assess the dynamic effects of transactions. More specifically, in his model the price effect of a transaction can last for more periods than the one period assumed implicitly in equation (5.7). Our regression based spread estimator can be extended easily to include more complex dynamics by adding lagged regressors to the regression models (5.9) and (5.9′). The parameters of interest are the coefficient of the current sign, trade size and inverted size, whereas the coefficients of the lagged variables are merely nuisance parameters. The parameter estimates using four (rather than one) lags of trade sign show only minor differences with the reported estimates and the conclusions do not change.

5 We do not impose restrictions on the coefficients of the lagged regressors. We do not want to run the risk of imposing invalid restrictions and thus misspecifying the model. Not imposing such restrictions does not affect the consistency of the estimators of the parameters of interest (S, α and γ).

6 Hausman et al. (1992) also censor trade size at the 99.5% quantile.

Table 5.8 Model based estimates of realised spread, with size effects.

model: $\Delta p_t = \beta_0 + \delta Q_t + \alpha z_t + \gamma/z_t + \text{lags} + \varepsilon_t$

	A. Paris (size censored at 2 NMS)				B. London (size censored at 5 NMS)			
	2δ	2α	2γ	Wald	2δ	2α	2γ	Wald
AC	0.181	0.059	2.399	18.513	0.976	-0.066	-7.459	4.123
	(23.338)	(2.870)	(4.279)		(8.018)	(-1.311)	(-1.817)	
AQ	0.145	0.014	3.979	52.172	1.283	0.002	3.167	0.241
	(36.234)	(1.745)	(6.944)		(11.391)	(0.048)	(0.486)	
BN	0.158	0.038	0.184	21.349	0.591	0.187	13.980	30.103
	(58.359)	(3.665)	(4.020)		(6.777)	(1.569)	(5.450)	
CA	0.172	0.012	0.107	8.364	0.819	-0.095	2.325	4.263
	(44.577)	(0.527)	(2.878)		(9.655)	(-1.501)	(1.370)	
CS	0.302	0.073	0.298	11.630	0.793	0.188	6.996	27.599
	(30.145)	(1.985)	(3.385)		(3.184)	(1.794)	(5.240)	
EX	0.115	0.017	0.168	15.473	0.737	0.037	1.724	3.261
	(41.387)	(1.163)	(3.931)		(10.204)	(0.572)	(1.770)	
OR	0.257	0.047	0.854	37.772	0.735	0.113	70.061	9.312
	(42.213)	(2.446)	(6.137)		(7.021)	(2.189)	(2.677)	
RI	0.283	0.008	0.965	4.947	0.680	0.036	21.844	1.943
	(20.015)	(0.226)	(2.077)		(3.696)	(0.408)	(1.381)	
SE	0.313	-0.056	0.561	9.337	2.163	-0.166	-13.606	3.896
	(29.231)	(-2.053)	(1.270)		(4.362)	(-1.081)	(-1.971)	
UAP	0.348	-0.000	3.764	77.760	0.814	0.012	0.606	0.534
	(30.332)	(-0.021)	(7.981)		(8.692)	(0.346)	(0.719)	

Notes: Wald is $\chi^2(2)$ test of joint significance of α and γ;
HAC t-ratios below parameter estimates; see also notes to Table 5.6.

The following tables show the bid-ask spread for different transaction sizes, estimated from the parameters of Table 5.8.

Estimated bid-ask spread Paris.

size	0.1	0.5	1.0	2.0
AC	0.199	0.213	0.241	0.299
AQ	0.155	0.154	0.160	0.173
BN	0.162	0.177	0.196	0.234
CA	0.173	0.178	0.184	0.196
CS	0.313	0.340	0.376	0.449
EX	0.117	0.123	0.132	0.148
OR	0.265	0.281	0.305	0.352
RI	0.293	0.289	0.292	0.299
SE	0.311	0.286	0.257	0.201
UAP	0.367	0.351	0.349	0.348

Estimated bid-ask spread London.

size	0.1	0.5	1.0	2.0	5.0
AC	0.932	0.935	0.906	0.841	0.644
AQ	1.289	1.285	1.286	1.288	1.294
BN	0.666	0.696	0.783	0.967	1.525
CA	0.818	0.774	0.726	0.631	0.347
CS	0.882	0.901	0.988	1.173	1.734
EX	0.747	0.757	0.775	0.811	0.921
OR	1.026	0.848	0.867	0.975	1.306
RI	0.902	0.741	0.738	0.763	0.865
SE	2.079	2.067	1.991	1.829	1.334
UAP	0.819	0.821	0.827	0.839	0.876

Note: size measured in NMS.

5.8 Summary and conclusions.

In this chapter we compared the cost of trading French shares in Paris and in London. The estimates of the average quoted spread, which reflect the cost of immediate trading, suggest that the Paris Bourse is cheaper than London's SEAQ International for small transactions, roughly up to the normal market size. For larger sizes, however, the Paris limit order book often does not contain enough limit orders and the average quoted spread rises steeply, hence the Paris market is not very deep. The London market with its competing market makers provides more liquidity at larger trade sizes. The quoted spread in London for small sizes is however relatively large.

The estimates of the realised spread show a slightly different picture. It appears that the few large transactions that are executed in Paris (often "crosses") have a fairly low spread, lower than the spread in London. Our regression-based estimates suggest that at trade sizes of twice NMS the realised spread is still considerably lower than in London. On the whole, we conclude that if the trader is patient and prepared to wait for counterparties, transaction cost for large sizes can be fairly low in Paris compared with SEAQ-International.

The full cost of trading on either exchange includes taxes and other levies as well. Information on such explicit transaction costs are presented in London Stock Exchange (1992a). The commissions and fees in London are on average 0.14% of the transaction value and in Paris about 0.5% (these percentages are for a large transaction of 1 million ECU, roughly FF 7 million). Thus explicit transaction costs are higher in Paris for large transactions. One reason is that in London many large deals are done on a "net" basis, i.e. commissions are included in the price.

A theoretically interesting result is that the realised spread is virtually flat in trade size. Hence, we do not confirm the predictions of the pure inventory control or adverse selection microstructure theory (that the spread should be an increasing function of trade size) except for the quoted spread (where the spread increases with trade size by construction). Our estimates of a simple model for transaction prices confirm this result and indicate mild support for the hypothesis that part of the order processing cost is fixed per transaction rather than per share.

Appendix. Adjustment for bias due to misreported transaction times.

As explained in the main text, the $S_R(z)$ estimates of the average realised spread in London for transaction sizes smaller than NMS are sometimes larger than the average quoted spread, $S_Q(z)$. This seems impossible, because the true realised spread has to be smaller than the quoted spread since market makers are obliged to provide the best quoted price for transactions smaller than NMS. This anomaly is probably explained by a timing bias due to misreported transaction times in London. In this appendix we propose a model for the impact of timing bias on estimates of the realised spread that can also be used to correct the S_R estimates for this bias.

Let S(z) be the average realised spread (as a function of size) that we would want to estimate. Suppose that the transaction is reported late, say at time t+k. In general the midprice recorded at time t+k is different from the mid-price at time t, so that in fact we estimate

$$(5A.1) \qquad S_R(z) = E|S(z) + x_t|$$

where x_t denotes the change in the mid-price in the interval between the time that the transaction actually took place and when it was reported. Suppose that x_t is normally distributed with mean 0 and variance σ^2. Then we can apply the expressions in Amemiya (1985, p. 367), who shows that for a normally distributed variable $y \sim N(\mu,\sigma^2)$, the conditional expectation of y, given $y>0$ is

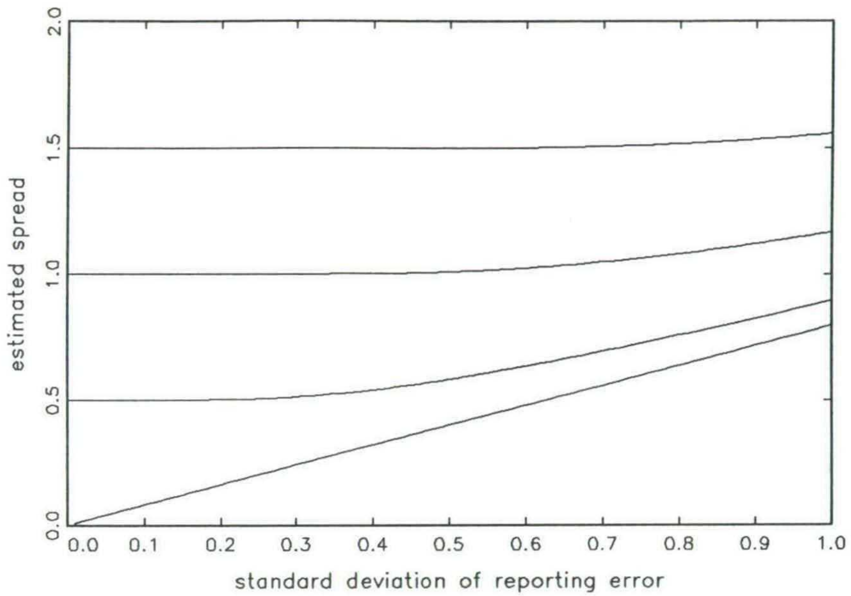
$$(5A.2) \qquad E(y|y>0) = \mu + \sigma \cdot \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}$$

where ϕ and Φ are the standard normal density and the cumulative standard normal distribution, respectively. Using this result the expectation of the absolute value in (5A.1) can be written as

$$(5A.3) \qquad S_R(z) = S(z)(2\Phi(a)-1) + 2\sigma\phi(a), \qquad a \equiv S(z)/\sigma$$

Figure 5.5 shows that the realised spread measure $S_R(z)$ is always larger than the S(z) that we want to estimate. The estimates reported in Table 5.6, panel B will therefore in general overstate the true spread if $\sigma>0$.

Figure 5.5: Timing bias for true spread 0, 0.5, 1 and 1.5



We now turn to a method to correct for timing bias. Fundamental to the correction is the assumption that the variance of the timing error, σ^2, is independent of the transaction size. Moreover, it is known that most small transactions are at the touch (London Stock Exchange (1992b)). Thus, for small transactions the quoted spread and the true realised spread should be the same: $S(z)=S_Q(z)$. In Table 5.5, panel B the average quoted spread by size class can be found. The first step in the correction procedure is to solve (5A.3) for σ, given $S(z)=S_Q(z)$ in the smallest size class. The second step is to compute $S(z)$ for all other size classes from (5A.3), given the estimate of σ obtained in the first step and the estimated values of $S_R(z)$ from Table 5.6, panel B.

There is one problem with the procedure outlined just before. If we take the smallest size class to be the class from 0 to 0.1 NMS, we estimate quite a large σ. In fact, the estimated σ is so large that (5A.3) sometimes does not have a solution. Therefore, we choose to base the estimate of σ on the average quoted spread for all transactions up to 1 NMS. Because quotes are firm up to 1 NMS, the quoted spread is an upper bound for the realised spread for this size class. Therefore, the estimated σ from solving (A.3) given $S(z)=S_Q(0,1)$ gives a lower bound for the true timing error. This estimate of σ will therefore yield a conservative correction of the realised spread for other size classes.

Chapter 6

Intra-day price effects of trading on the Paris Bourse

6.1 Introduction.

In this chapter we analyse the dynamic price effects of trading on the Paris Bourse. The issue is of interest both for practitioners and theorists. Traders, especially of large 'blocks' of shares, may be worried that their transactions might put a downward or upward pressure on future stock prices. A closely related aspect of the quality of a securities market is 'resiliency', i.e. the speed at which the stock price reverts to its equilibrium value after a large trade. The issue is also of interest for the theory of stock market micro structure, because it allows us to shed some light on the main theoretical explanations for the price effect of transactions: inventory control by market makers and adverse selection due to asymmetric information.

The inventory control literature, e.g. the paper by Ho and Stoll (1981), usually considers a dealership market where a professional market maker provides liquidity and absorbs order imbalances. All trades are assumed to be executed through the market maker, whose position in the stock (inventory) in general deviates from his desired inventory level. Therefore, the market maker poses bid and ask quotes to encourage trades that will restore his inventory to the desired level. For example, starting out from a balanced inventory, after a large sale to the market maker (at the bid), inventory will be higher than desired, and the market maker will revise his quotes downward to encourage purchases by the public. Hence, the trade moves the price in the direction of the trade.

The second explanation that is offered in the literature for price effects of trading supposes that there is asymmetric information between the market maker and the public, because some traders might have superior information on the true value of the stock. Traders with superior information will have an incentive to trade on one side of the market only. For example, a trader who knows that the quoted price is lower than the true value will have an incentive to buy. A trade at the ask, initiated by a buyer, therefore gives the market maker a signal that his quoted prices are too low, and he will revise his quotes upward. Moreover, informed traders have an incentive to trade large quantities, in order to maximise the benefits of trading on perishable asymmetric information, cf. Glosten and Harris (1988) and Easley and O'Hara (1987). Therefore, there is also adverse selection in the size of the transaction: large trades are more likely to be initiated by informed traders than small trades.

The market structure of the Bourse (discussed in detail in Chapter 5) differs considerably from a dealership market. There are no specialists or professional market makers. Instead, liquidity is provided by the public limit order book. In line with most research on the micro structure of securities markets (especially most papers that analyse NYSE data), we assume that microstructure theories developed for dealership markets also apply to the continuous auction market of the Paris Bourse. Another motivation for using these theories is that the issuers of limit orders generally face the same decision problems as market makers, and therefore require similar compensation for their services.

In this chapter we discuss and apply various methods to measure the price effects of trading. Estimates of the price effects of inventory control and asymmetric information allow us to split the quoted bid-ask spread into components related to these effects and a third component due to the pure processing cost of a transaction. In Section 2, we analyse the price effects of trading using the structural model of Madhavan and Smidt (1991). We slightly extend the model to take the dependence of the processing cost on the size of transactions into account. It appears that the inventory control component of the spread is not identified in the Madhavan and Smidt model if the inventory level cannot be observed. In the next sections we try to solve this problem. We measure the adverse selection and inventory control costs by the price effect of a transaction on subsequent quote mid-points or transaction prices. The key identifying assumption is that adverse selection causes a permanent shift in prices, whereas the other spread components have only temporary price effects. In Section 3 the price effects of trading are estimated assuming that the trading process is exogenous. In Section 4 we develop a model for simultaneous determination of prices and trades. Section 5 reports the empirical results with the simultaneous model. In Section 6 we conclude.

6.2 Structural models of transaction prices.

In this section we describe some structural models for transactions data. We propose an extension to the model of Madhavan and Smidt (1991) to allow for dependence of processing cost on trade size. It is demonstrated that several other microstructure models can be written as special cases of the extended Madhavan and Smidt model. One special case, the Glosten and Milgrom (1988) model, is estimated on Paris Bourse data.

6.2.1 The Madhavan and Smidt model.

Madhavan and Smidt (1991), hereafter referred to as MS, present an elegant model for transaction prices. In the model, the market maker poses bid and ask quotes based on his best guess of the true value of the stock and on his current inventory. His expectation of the value of the stock after a transaction is updated using Bayes' rule. Define the following variables:

p_t = transaction price
Q_t = 'sign' of the transaction[1]
z_t = signed size (number of shares traded)
I_t = inventory of market maker *before* the transaction
y_t = expected value of the stock *before* the transaction
μ_t = expected value of the stock *after* the transaction
e_t = publicly observed change in the value of the stock

The first equation of the Madhavan and Smidt model states that the transaction price is equal to the expected value of the stock *after* the transaction, minus a discount if inventory is positive, plus a fixed 'processing cost':

(6.1) $$p_t = \mu_t - \gamma I_t + \psi Q_t.$$

The expected value after the transaction, μ_t, is based on all publicly available information after the transaction, and includes the adverse selection price effect; inventory control is modelled through the γI_t term. In the Bayesian updating scheme of expectations assumed by MS, the 'posterior' mean μ_t is obtained by taking a weighted average of the 'prior' mean (the expected value before the transaction) and the 'excess demand' for shares

(6.2) $$\mu_t = \pi y_t + (1-\pi)[p_t+z_t/\alpha].$$

[1] The *sign* is defined as follows: +1 if the transaction is initiated by the buyer (at the ask) and -1 if the transaction is initiated by the seller (at the bid).

The parameter π reflects the degree of information asymmetry between the market maker and the public; if there is no asymmetry then $\pi=1$. The parameter α is related to the slope of the demand function for shares. To close the model, it is supposed that the expected value before the transaction, y_t, is equal to the expected value after the previous transaction plus the publicly observed rise in the value of the shares between the two transactions

$$(6.3) \qquad y_t = \mu_{t-1} + e_t.$$

Together, (6.1)-(6.3) constitute the model. Substituting (6.2) in (6.1), one obtains the transaction price as a function of the expected value before the transaction, minus the discount for inventory, plus the price effects of a transaction

$$(6.4) \qquad p_t = y_t - (\gamma/\pi)I_t + \lambda z_t + (\psi/\pi)Q_t,$$

where $\lambda \equiv (1-\pi)/\pi\alpha$. Thus, the transaction price is equal to the quote midpoint, $y_t-(\gamma/\pi)I_t$, plus or minus half the quoted bid-ask spread for that particular size, $\lambda z_t+(\psi/\pi)Q_t$. Note that the quoted bid-ask spread does not depend on the cost of inventory control. This feature of the model, which was introduced implicitly in equation (6.1), is rather counter-intuitive, and we shall relax this assumption in the next sub-section.

Substituting y_t out of (6.2) using (6.3), and then substituting $\mu_t=p_t+\gamma I_t-\psi Q_t$ from (6.1), yields an equation in observables only:

$$(6.5) \qquad \Delta p_t = \lambda z_t - (\gamma/\pi)I_t + \gamma I_{t-1} + (\psi/\pi)Q_t - \psi Q_{t-1} + e_t.$$

This is the equation that MS estimate[2] on a sample of NYSE transaction price series. Their estimated coefficients of z_t and Q_t are often significantly positive, which supports the presence of both adverse selection and processing costs. The estimates of the weight of prior information, π, based on the ratio of the coefficients of Q_t and Q_{t-1}, lie around 2/3. On the other hand, the estimated coefficients of I_t and I_{t-1} are usually not significant and always nearly equal, but of different sign, which would imply a value of π close to 1.

---

2 Except for a constant term that is included if the desired inventory level or the expected value of e_t are non-zero.

A special case of the MS model arises when there is no asymmetric information, so that $\pi=1$, and all trades are executed via the market maker, so that $I_t=I_{t-1}-z_{t-1}$. In that case, the model reduces to

$$\text{(6.6)} \qquad \Delta p_t = \gamma z_{t-1} + \psi(Q_t-Q_{t-1}) + e_t,$$

which is precisely the Ho and Macris (1984) inventory model. In this model the quoted spread is constant (2ψ) and does not depend on trade size. Ho and Macris justify this assumption by assuming that all trades are of the same size. The cost of carrying inventory is then included in the coefficient ψ. Note also that inventory data are not required to estimate the Ho-Macris model.

6.2.2 An extension to size-dependent inventory control and processing costs.

The assumption in (6.1) that the transitory cost components (inventory control and processing costs) do not depend on trade size is clearly unrealistic. Therefore, we propose to augment the price equation (6.1) with processing cost and inventory control costs that depend on trade size.

The first change to (6.1) is a term βz_t that reflects the part of the processing cost that depends on trade size, as in the model of Glosten and Harris (1988) (which does not have an inventory control component, however).

The model of Ho and Stoll (1981), which is a pure inventory control model, suggests that the proportional part of the inventory control cost is equal to γz_t, where γ is the coefficient of the inventory level in the pricing schedule. Stoll (1989) provides an intuitive motivation for this: a transaction of size z_t moves the mid-quote by $-\gamma$ times the change in inventory, ΔI_{t+1}. Therefore, using that $\Delta I_{t+1}=-z_t$, the 'inventory control' price effect of a transaction is γz_t.

The third extension is a random pricing error, u_t, that captures other influences on the transaction price, such as price discreteness and omitted variables, i.e. other factors that influence the spread, but are not modelled. We assume that u_t is uncorrelated with the other variables in the price equation.

With these changes to (6.1), the pricing equation becomes

$$\text{(6.7)} \qquad p_t = \mu_t - \gamma I_t + (\beta+\gamma)z_t + \psi Q_t + u_t,$$

The informational structure of the model is not changed and the updating equations (6.2) and (6.3) are retained. Hence, the transaction prices in the augmented model are determined by

$$(6.8) \qquad p_t = y_t - (\gamma/\pi)I_t + (\lambda + \beta/\pi + \gamma/\pi)z_t + (\psi/\pi)Q_t + u_t.$$

The quote mid-point is $y_t-(\gamma/\pi)I_t$ and the quoted spread for size $|z|$ is

$$S(z) = 2(\psi/\pi + (\lambda + \beta/\pi + \gamma/\pi)|z|)$$

The decomposition of the half-spread into cost components is as follows:

Processing cost	$\psi + \beta\|z\|$
Inventory control cost	$\gamma\|z\|$
Adverse selection cost	$\psi^* + (\lambda + \beta^* + \gamma^*)\|z\|$

where $\psi^*=(\frac{1-\pi}{\pi})\psi$, $\beta^*=(\frac{1-\pi}{\pi})\beta$, and $\gamma^*=(\frac{1-\pi}{\pi})\gamma$. From this decomposition it is clear that there is interaction between adverse selection and the other cost components: the presence of the other cost components strengthens the adverse selection effect. Glosten (1987) provides an intuitive argument for this: if there are other transaction costs besides adverse selection costs, noise traders are less likely to trade (compared with the situation where there are no other costs) and therefore transactions that do take place give stronger signals about the true value of the shares.

Similar substitutions as before yield the following regression equation:

$$(6.9) \qquad \Delta p_t = (\lambda+(\beta+\gamma)/\pi)z_t - (\beta+\gamma)z_{t-1}$$
$$- (\gamma/\pi)I_t + \gamma I_{t-1} + (\psi/\pi)Q_t - \psi Q_{t-1} + \xi_t,$$

where $\xi_t \equiv e_t+\Delta u_t$. If data on the inventory level are available, this equation can be estimated and all structural parameters are identified.

If all trades are executed through a market maker, the size and inventory variables are multicollinear because $I_t=I_{t-1}-z_{t-1}$. Upon substitution of this identity, an equation similar to (6.5), the regression equation of MS, is obtained

$$(6.10) \qquad \Delta p_t = (\lambda+\beta/\pi+\gamma/\pi)z_t + (\beta-\gamma^*)I_t - \beta I_{t-1} + (\psi/\pi)Q_t - \psi Q_{t-1} + \xi_t.$$

Compared with (6.5), the coefficient of the trade size z_t has changed; it now depends on all cost components. The coefficients of the inventory levels also have changed; both depend on the proportional part of the processing cost, β. If $\gamma=0$, the coefficients of current and lagged inventory are equal except for their sign (β and $-\beta$, respectively), which is what MS found empirically. Probably the hypothesis that $\gamma=0$ cannot be rejected for the MS estimates. Our conclusion is that the evidence of MS for the inventory effect is very weak; processing costs proportional to trade size seem more likely to be empirically relevant.

Other evidence for the inventory control effect is at best weak. Madhavan and Smidt (1992) and Hasbrouck and Sofianos (1992) estimate inventory control models on samples of data from the NYSE. In particular, they test for mean reversion in the inventory levels of specialists. Both papers are only able to find mean reversion in the inventory levels if they allow for speculative shifts in the desired inventory level. Moreover, the estimated reversion to the desired level is very slow, and takes a number of days. Therefore, for analysing intra-day price effects, inventory control is perhaps not so relevant.

Another problem with the inventory effect is that estimation of equations (6.9) or (6.10) requires data on the inventory level of market makers. For the Paris Bourse, such data are not directly available. We also cannot construct inventory levels because we do not observe all transactions. For example, we do not know who trades in London or on the inter-dealer market. Therefore, we assume that there is no inventory effect[3] ($\gamma=0$). In this case, equation (6.9) takes a particularly simple form:

$$(6.11) \qquad \Delta p_t = (\lambda+\beta/\pi)z_t - \beta z_{t-1} + (\psi/\pi)Q_t - \psi Q_{t-1} + \xi_t.$$

The components of half the spread for size $|z|$ in this model are

Processing cost	$\psi + \beta\vert z\vert$
Adverse selection cost	$\psi^* + (\lambda + \beta^*)\vert z\vert$

In fact, this is the Glosten and Harris (1988) model, where both adverse selection and processing cost components have a constant, size-independent part and a part

3 Alternatively, we could assume that there is no asymmetric information, so that $\pi=1$. From the estimates it will become clear that this hypothesis is rejected by the data.

that varies with trade size. The components of the spread in the Glosten-Harris model can be estimated directly by the following re-parametrisation of (6.11)

$$\Delta p_t = (\lambda+\beta^*)z_t + \beta\Delta z_t + \psi^*Q_t + \psi\Delta Q_t + \xi_t. \tag{6.12}$$

The permanent effects, attributed to adverse selection, are measured by the coefficients of the level variables, whereas the transitory processing cost is measured by the coefficients of the variables in first differences.

A number of econometric issues concerning the estimation of equations (6.9) to (6.12) require special attention. Following Harris (1987) and Hasbrouck (1991b), who argue that observed covariance patterns in transaction returns are more consistent with transaction time than with calendar time, we assume that the relevant 'clock' is transaction time. Contrary to Hasbrouck (1993), we include a constant term in the model to capture the average return between transactions (i.e. a non-zero mean of e_t). The *variance* of the errors is unspecified by the model. For several reasons, it is likely that the errors are heteroskedastic. For example, the variance may depend on the time of day, and the variance may depend on the trade size. Moreover, due to the presence of the pricing error u_t, ξ_t has an MA(1) serial correlation pattern. With this error structure, OLS gives consistent point estimates, but the usual standard error formula is incorrect. We compute heteroskedasticity and autocorrelation consistent (HAC) standard errors by the method of Newey and West (1987)[4].

6.2.3 Data and empirical results.

Our data set consists of a complete record of transactions on the Paris Bourse in a two month period in the summer of 1991. In this chapter, we use the same data as in Chapter 5. Recall that these data concern time series of limit order prices and transactions for ten large firms; the data are described in detail in Section 5.3. In this chapter we analyse the dynamic effects of transactions on the Paris Bourse only. French shares are also traded on other exchanges, especially London's SEAQ International. It would appear natural to merge the transaction files of both exchanges to one large data set. There are good arguments not to include the London data in the analysis, however. Transactions in London are negotiated between traders and market makers over the phone, and are often reported to other traders with a considerable time lag or not at all. The trading process in London

[4] We also estimated the model by a more efficient GLS estimator that takes account of the MA(1) error structure. These estimates were very close to the OLS estimates.

is hardly visible for outsiders; the only thing that is publicly observed are the market makers' quotes, but these are often adjusted slowly and do not give a good indication of actual transaction prices (see Section 5.7). Therefore, we decided to analyse only the transactions from the Paris Bourse.

In Table 6.1 the estimates of the Glosten-Harris model (6.12) are presented.

Table 6.1 Estimates of the Glosten-Harris model (6.12).

$$\Delta p_t = \kappa + (\lambda+\beta^*) z_t + \beta \Delta z_t + \psi^* Q_t + \psi \Delta Q_t + \xi_t$$

	κ	$\lambda+\beta^*$	β	ψ^*	ψ	π	S(0)	S(1)
AC	0.0036	-0.0219	-0.0040	0.0302	0.0816	0.73	0.2234	0.1876
	[2.53]	[1.65]	[0.35]	[11.26]	[31.22]		(0.27)	(0.09)
AQ	0.0046	0.0157	-0.0227	0.0208	0.0590	0.74	0.1595	0.1453
	[6.60]	[3.51]	[3.79]	[15.32]	[43.19]		(0.26)	(0.50)
BN	0.0045	0.0196	-0.0103	0.0163	0.0653	0.80	0.1631	0.1817
	[7.57]	[3.52]	[2.03]	[14.84]	[62.62]		(0.20)	(0.40)
CA	0.0015	0.0211	-0.0212	0.0181	0.0779	0.83	0.1919	0.1919
	[1.80]	[4.04]	[4.54]	[11.40]	[46.31]		(0.19)	(0.41)
CS	0.0152	0.0255	0.0059	0.0425	0.1227	0.74	0.3304	0.3932
	[5.45]	[1.52]	[0.42]	[9.66]	[30.19]		(0.26)	(0.35)
EX	0.0029	0.0079	-0.0101	0.0179	0.0496	0.74	0.1350	0.1306
	[4.38]	[2.13]	[2.98]	[13.08]	[36.36]		(0.27)	(0.40)
OR	0.0068	0.0011	-0.0139	0.0344	0.1051	0.75	0.2790	0.2533
	[5.16]	[0.08]	[1.19]	[13.39]	[41.26]		(0.25)	(0.28)
RI	0.0135	-0.0027	-0.0311	0.0484	0.1195	0.71	0.3358	0.2681
	[4.56]	[0.13]	[1.59]	[9.69]	[23.95]		(0.29)	(0.34)
SE	0.0029	-0.0048	-0.0254	0.0370	0.1178	0.76	0.3097	0.2495
	[1.61]	[0.30]	[1.85]	[11.55]	[33.80]		(0.24)	(0.26)
UAP	0.0120	-0.0083	-0.0431	0.0642	0.1672	0.74	0.4629	0.3600
	[4.93]	[0.60]	[3.39]	[13.15]	[33.16]		(0.28)	(0.31)

Notes: dependent variable is percentage change in transaction price;
model estimated by OLS, overnight returns omitted; size (z) censored at 2 NMS;
sample 24-5-91 to 31-6-91. Newey-West (1987) t-values in square brackets.
S(0) is the spread for z=0, S(1) is spread for z=NMS with the fraction of S attributed to adverse selection in parentheses.

Like in Chapter 5, we excluded overnight returns and opening prices from the estimation. The size of the transaction is censored at 2 times NMS, which corresponds roughly to the 99.5 percentile of the size distribution, see Tables 5.1 and 5.2, so that only about 5 out of every 1000 transactions are affected. The reason for censoring is to mitigate the effect of very large trades on the estimates, see also Hausman, Lo and MacKinlay (1992). In the estimation, the dependent variable is the percentage change in transaction prices[5], so that the parameter estimates can be interpreted as relative (percentage) price effects.

The table shows that the fixed processing cost per transaction, ψ, is always highly significant, but varies quite a lot, from 0.06% for Elf-Aquitaine to 0.17% for UAP. The proportional processing cost, β, is negative in all cases but one (Axa-Midi) where it is positive but insignificant. The highly significant estimates of ψ^* indicate that the adverse selection cost has a fixed, size-independent component for all series. This result strongly contradicts the findings of Glosten and Harris (1988), who report no significant fixed adverse selection cost[6]. Only in four cases the proportional part of adverse selection cost is significantly positive. In the other cases the proportional adverse selection cost is not significant. The weight of the prior mid-price in the new mid-price, π, is estimated close to 0.75 for all series, which is slightly higher than the average of 2/3 reported by Madhavan and Smidt (1991).

Together, the processing cost and adverse selection cost make up the bid-ask spread. Table 6.1 also shows estimates of the bid-ask spread for a very small transaction and a transaction at NMS, which is quite a large transaction by Paris standards[7]. The estimates suggest that usually the spread is declining in trade size, so that the declining processing cost dominates the increasing adverse selection cost. In only two cases the spread increases with trade size. The estimated spreads are comparable to the results obtained in Table 5.8. Table 6.1 also reports the proportion of the spread that can be attributed to adverse selection. For the smallest size, about 25% of the spread is due to adverse selection. Because of the decreasing processing cost and increasing adverse selection cost, the adverse selection component at NMS is usually much larger, about 40% of the spread.

6.3 Reduced form estimation of price effects of transactions.

In the structural model of the previous section, the inventory control cost is not identified if no data on the inventory level are available. Moreover, the estimates of the asymmetric information cost component use the ad hoc assumption that the

5 $\Delta p_t = 100 \cdot \ln(P_t/P_{t-1})$, where P_t is the observed transaction price.
6 In their empirical work, Glosten and Harris found that only the fixed part of the processing cost and the part of the adverse selection cost that is proportional to trade size were significant.
7 Table 5.2 shows that 1 NMS is about the 99th percentile of the transaction size distribution on the Bourse.

price effect of a transaction is fully revealed in the prices after one transaction. An alternative method to estimate the adverse selection cost is by the expected long run price effect of a transaction. If the asymmetric information becomes public immediately after the transaction, the initial price effect is equal to the sum of the adverse selection effect and the inventory control effect. These observations allow us to base estimates of the adverse selection cost and inventory control cost components of the spread on the estimated price effects of trading.

In this section, the long run price effects are computed under the assumption that the transaction sign is exogenous, but possibly correlated over time. In order to estimate the spread and the initial price effect of a transaction we use the model of Stoll (1989) and George et al. (1991), which was also used in Section 5.7. The pricing equation of the model is

$$(6.13) \qquad p_t = y_t + (S/2)Q_t + u_t.$$

The transaction size is not included in the pricing equation; we could think of S as the spread for a transaction of average size[8]. The mid-quote y_t now includes the price discount due to inventory holdings.

If there is asymmetric information or inventory control, the market maker will revise his mid-quotes after a transaction. Let δ be the fraction of the spread by which the mid-quote initially is revised

$$(6.14) \qquad m_t = y_t + \delta S \cdot Q_t.$$

Between two trades public information on the stocks value may come in and therefore the mid-quote prior to the subsequent trade is a revision of the current mid-quote

$$(6.15) \qquad y_{t+1} = m_t + e_{t+1}.$$

The processing cost component of the spread can be estimated directly by a regression model. Combining (6.13)-(6.15), the transaction price changes can be expressed as

8 Suppose the true pricing equation were $p_t=y_t+aQ_t+bz_t$. This equation can be written as $p_t = y_t + (a+bE|z|)Q_t + b(z_t-E|z|Q_t) = y_t + (S/2)Q_t + u_t$, where u_t is by construction uncorrelated with Q_t.

(6.16) $\Delta p_t = \psi^* Q_t + \psi \Delta Q_t + \xi_t,$

where $\psi^* = \delta S$, $\psi = (1-2\delta)(S/2)$ and $\xi_t = e_t + \Delta u_t$. This form gives direct estimates of the processing cost, ψ, and the initial price effect, ψ^*, of a transaction of average size. An estimate of δ is easily found as a function of ψ and ψ^*.

So far, the model is a simplified version of the Madhavan and Smidt model[9]. New in this section is that we split up the initial price effect δS into components due to asymmetric information and inventory control by using the long run price effect as a measure of the adverse selection cost. Following Hasbrouck (1988) and Stoll (1989), we compute the long run price effect under the assumption that the process generating the transaction sign is exogenous to the price process.

There is some debate in the theoretical literature about the effect of trading on future transaction signs. In the presence of inventory control, there is an expected reversal of the sign of the transaction, i.e. after a buy we expect that the probability of the subsequent transaction being a sale increases. This effect is caused by efforts of the market maker to rebalance his inventory to the desired level. The effect of asymmetric information on the bid-ask bounce is less clear, however. Glosten and Milgrom (1985) merely assume that there is no serial correlation in the sign, but give no explicit reason for that. Stoll (1989) and Madhavan and Smidt (1991) assume that all private information is revealed to the public immediately after the transaction, and that therefore there is no effect on future trades, except for the shift in expected value. In other models, e.g. Kyle (1985) and Easley and O'Hara (1987), it takes longer for all private information to become public. In such a case the informed traders will keep trading on the same side of the market for a while in order to maximise the profits from having superior information. If so, the trade sign shows positive serial correlation. This is completely the opposite of the reversal of sign predicted by inventory control models.

Stoll (1989) assumes that there is a fixed probability (denoted by π) of a trade *reversal*, i.e. the probability that the next transaction is of opposite sign. In our opinion, an important shortcoming of Stoll's analysis is his definition of the inventory control component. Stoll's definition takes only takes the price effect one period ahead into account, where a definition that takes account of the complete future of trading seems more appropriate, because it usually takes more than one transaction to rebalance the market maker's inventory. Instead, we adopt

[9] With $\lambda = \gamma = \beta = 0$, $\psi = (1-2\delta)(S/2)$, $\psi^* = \delta S$, $S = 2(\psi + \psi^*)$ and $\pi = 1-2\delta$.

the approach of Hasbrouck (1988) by allowing for a more general serial correlation pattern in the trade sign Q_t.

As explained in the introduction, the adverse selection price effect can be measured by the long run price effect of a transaction. The initial price effect is a mixture of both inventory control and adverse selection effects. We use the following measure of the effect of a transaction on the mid-quotes after T trades:

$$(6.17) \qquad dp_T(q) = E(y_{t+T}-y_t \mid Q_t=q) - E(y_{t+T}-y_t).$$

In words, dp_T is the expected mid-quote change after T transactions, given the sign of the current transaction, minus the unconditionally expected mid-quote change. To compute $dp_T(q)$ in the model, rewrite (6.17) as follows:

$$(6.18) \qquad dp_T(q) = E\left(\sum_{i=1}^{T} \Delta y_{t+i} \mid Q_t=q\right) - E\left(\sum_{i=1}^{T} \Delta y_{t+i}\right)$$

Substituting $\Delta y_{t+1}=\delta S \cdot Q_t+e_{t+1}$ (derived from (6.14) and (6.15)) in (6.18) yields

$$(6.19) \qquad dp_T(q) = \delta S \sum_{k=0}^{T-1} \left(E(Q_{t+k} \mid Q_t=q) - E(Q_{t+k})\right),$$

where we assumed that the price changes between future transactions do not depend on the sign Q_t, $E_t(e_{t+i} \mid Q_t)=E_t(e_{t+i})$, so that the terms including e_{t+i} cancel. In the analysis, transactions at the bid and at the ask are treated symmetrically[10]. Therefore, our summary measure of the price effects of trading is

$$(6.20) \qquad dp_T \equiv E[Q_t dp_T(Q_t)] = \delta S \sum_{k=0}^{T-1} \gamma_Q^k$$

where γ_Q^k denotes the autocorrelation of order k of Q_t.

The long run price effect, which measures the asymmetric information effect, is $\lim_{T\to\infty} dp_T = \sum_{k=0}^{\infty} \gamma_Q^k$. Hence, the components of the spread are

10 Several authors, e.g. Holthausen et al. (1990), Keim and Madhavan (1992) and Chan and Lakonishok (1993), report that the price response to buyer and seller initiated transactions on the US stock market is asymmetric. We assume symmetry here to facilitate comparison of the results with the results of parametric models, where the asymmetry is not easily included.

Processing cost	$1-2\delta$
Adverse selection cost	$2\delta\cdot\sum_{k=0}^{\infty}\gamma_Q^k$
Inventory control cost	$2\delta\cdot(1-\sum_{k=0}^{\infty}\gamma_Q^k)$

The inventory control component is positive if and only if there is reversal of the sign: after a buy the cumulative number of expected sells must be larger than the cumulative number of expected buys ($\sum_{k=0}^{\infty}\gamma_Q^k<1$, or $\sum_{k=1}^{\infty}\gamma_Q^k<0$).

The estimates of equation (6.16) are reported in Table 6.2.

Table 6.2 Components of the bid-ask spread with exogenous trading.

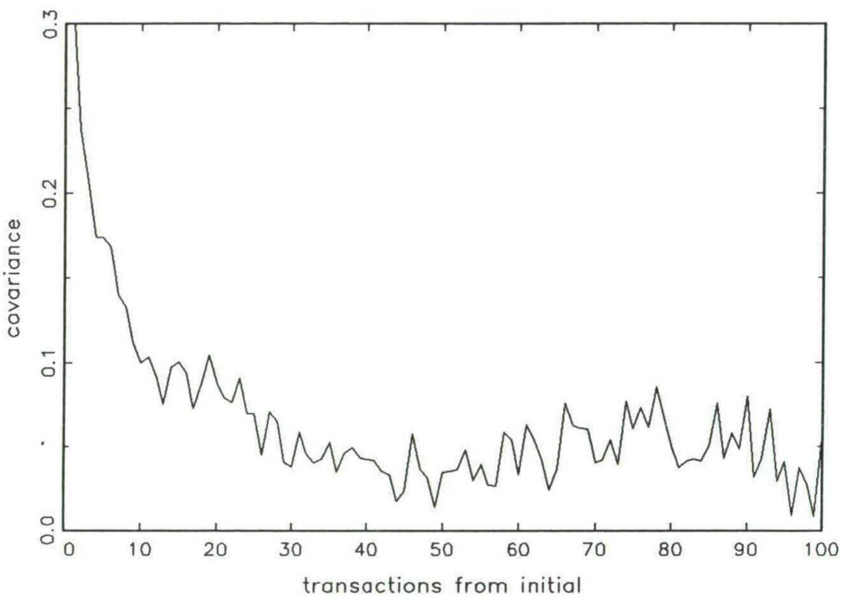
model: $\Delta p_t = \kappa + \psi^* Q_t + \psi\Delta Q_t + \xi_t$

	ψ^*	ψ	S	γ	PC	AI	IC
AC	0.0272	0.0822	0.2189	0.2527	0.75	0.56	-0.31
	[13.71]	[37.57]					
AQ	0.0237	0.0549	0.1572	0.2275	0.70	0.60	-0.30
	[22.82]	[47.11]					
BN	0.0183	0.0642	0.1649	0.1518	0.78	0.34	-0.11
	[19.91]	[68.52]					
CA	0.0211	0.0740	0.1923	0.3303	0.77	0.74	-0.51
	[18.54]	[50.88]					
CS	0.0465	0.1233	0.3397	0.1878	0.73	0.44	-0.16
	[13.34]	[35.09]					
EX	0.0198	0.0474	0.1345	0.2215	0.71	0.61	-0.31
	[19.83]	[41.34]					
OR	0.0345	0.1038	0.2766	0.1812	0.75	0.41	-0.16
	[17.01]	[46.51]					
RI	0.0478	0.1159	0.3275	0.1931	0.71	0.40	-0.10
	[11.91]	[26.69]					
SE	0.0364	0.1146	0.3020	0.2568	0.76	0.40	-0.16
	[13.93]	[37.72]					
UAP	0.0622	0.1519	0.4427	0.2348	0.72	0.53	-0.25
	[17.11]	[37.09]					

Notes: see also Table 6.1; $S=2(\psi^*+\psi)$ is the estimated average bid-ask spread.
γ is first order autocorrelation of sign. Sum of auto-covariances of sign truncated at K=20.
Components of the spread: PC: processing cost; AS: adverse selection cost; IC: inventory control cost.

As before, the dependent variable is the percentage transaction price change. The estimated spread and initial price effects for a transaction of average size are similar to those estimated by the Glosten-Harris model in Table 6.1. The numerical breakdown of the components of the spread yields very counter-intuitive results. Because the sum of the autocorrelations $\sum_{k=1}^{\infty}\gamma_Q^k$ is always positive, we obtain *negative* estimates of the inventory control component. This is illustrated in Figure 6.1, where the autocorrelations of the trade sign for the Accor series are graphed. The most striking result is the very strong positive serial correlation at low lags, which is in sharp contradiction with the assumptions of Stoll (1989) that the sign immediately reverses. Even in the long run there seems to be no reversal, all covariances are positive, which casts doubt on the existence of an inventory control effect for this series. The pattern for the other series is very similar. Hasbrouck (1988) reports similar results for NYSE stocks.

Figure 6.1: Autocorrelation function of sign Accor



Summarising, the results cast doubt both on the existence of an inventory control effect and on the assumption that information carried by trades is immediately disseminated to the market after a transaction. We refrain from possible explanations of this result until the concluding section.

6.4 A simultaneous model for prices and transactions.

In this section, we develop measures for the price effects of a transaction that do not need the assumption that the trading process is exogenous.

6.4.1 Non-parametric measures of price effects.

In the previous section we already defined a measure (dp_T) for the price effects of trading. It is easy to prove that dp_T is equal to the covariance between $y_{t+T}-y_t$ and Q_t:

$$
\begin{aligned}
(6.21) \qquad dp_T &\equiv E[dp_T(Q_t)Q_t] \\
&= E\left(E(y_{t+T}-y_t \mid Q_t=q)Q_t - E(y_{t+T}-y_t)Q_t\right) \\
&= E\left((y_{t+T}-y_t)Q_t\right) - E(y_{t+T}-y_t)E(Q_t) \\
&\equiv Cov(y_{t+T}-y_t, Q_t).
\end{aligned}
$$

This definition of the price effects of trading needs no assumptions on the joint dynamic process of price changes and sign (except covariance stationarity).

It is difficult, however, to extend this non-parametric measure to more explanatory variables. For example, the transaction size, which we found in Section 6.2 to be an important determinant of the bid-ask spread, is a continuous variable. Non-parametric measures conditional on size can only be computed if we categorise the data or use kernel-type estimators. Another problem is that the price effect probably also depends on the past (lagged values of price, sign and size), which requires high dimensional non-parametric measures, which are much harder to compute. In this section, a parametric simultaneous dynamic model of price changes and transaction characteristics, based on the work of Hasbrouck (1991a,b), is used to analyse the price effects of trading conditional on five lags of returns, sign and size.

6.4.2 Framework for dynamic analysis of prices.

To analyse the dynamic effects of processing costs, inventory control and asymmetric information, consider the following decomposition of the stock price, due to Hasbrouck (1993):

(6.22) $$p_t = \mu_t + s_t.$$

The first part, μ_t, is the semi-strong form 'efficient price' of the stock, i.e. μ_t is the best guess of the stocks value given all publicly available information. The second part, s_t, is a disturbance term that measures the deviation of the transaction prices from the efficient price. These deviations are assumed to be transitory, hence s_t is (weak-sense) *stationary*, so that in the absence of new shocks the transaction prices eventually converge to the efficient price, μ_t. Clearly, in the presence of asymmetric information trading affects μ_t. Because of their transitory nature, the processing cost and inventory control effect belong to the stationary part, s_t.

The price, being the sum of a non-stationary and a stationary term, is itself non-stationary. The empirical model is therefore specified on the first differences of prices (returns, $r_t=\Delta p_t$). The returns can be split up in the uncorrelated efficient price change ($w_t=\Delta\mu_t$) and an overdifferenced part:

(6.23) $$r_t = w_t + \Delta s_t.$$

In general, w_t is uniquely identified, but different identifying assumptions for s_t are possible, see Hasbrouck (1993). We use a vector auto regressive (VAR) model to identify w_t and s_t, which corresponds to the Beveridge and Nelson (1981) identification.

6.4.3 A parametric model of price effects.

The building blocks of the VAR model are dynamic equations for returns and trade attributes, such as sign and size. Returns are regressed on their own past and current and lagged values of a vector of trade attributes, x_t:

(6.24) $$r_t = a(L)r_{t-1} + b_0x_t + b(L)x_{t-1} + e_{1t},$$

where a(L) and b(L) are conformable vectors of lag polynomials. The coefficient vector b_0 measures the instantaneous price effect of a transaction. The trade attributes x_t are modelled by an autoregressive model with possible feedback from previous returns

(6.25) $$x_t = c(L)r_{t-1} + d(L)x_{t-1} + e_{2t}.$$

Because x_t is included in the equation for r_t, we impose the identification restriction that the innovations e_{1t} and e_{2t} are uncorrelated. Together, (6.24) and (6.25) constitute a bivariate Vector Auto Regression (VAR):

$$(6.26)\qquad \begin{pmatrix}1 & -b_0\\ 0 & I\end{pmatrix}\begin{pmatrix}r_t\\ x_t\end{pmatrix} = \begin{pmatrix}a(L) & b(L)\\ c(L) & d(L)\end{pmatrix}\begin{pmatrix}r_{t-1}\\ x_{t-1}\end{pmatrix} + \begin{pmatrix}e_{1t}\\ e_{2t}\end{pmatrix},\quad V\begin{pmatrix}e_{1t}\\ e_{2t}\end{pmatrix} = \begin{pmatrix}\sigma^2 & 0\\ 0 & \Omega\end{pmatrix}.$$

This model allows a very general dependence of price changes and trade sign (and size) on the past, without the restriction that x_t is exogenous.

The effects of shocks on future returns and other variables are more easily found by inverting the VAR to a Vector Moving Average (VMA) representation[11]:

$$(6.27)\qquad \begin{pmatrix}r_t\\ x_t\end{pmatrix} = \begin{pmatrix}a^*(L) & b^*(L)\\ c^*(L) & d^*(L)\end{pmatrix}\begin{pmatrix}e_{1t}\\ e_{2t}\end{pmatrix}.$$

The coefficients of the lag polynomials in (6.27) are the effects of past shocks to the current return and trade attributes. For example, b^*_τ is the effect of $e_{2,t-\tau}$ on r_t. The effects of a shock on the price level can be obtained by taking partial sums of the coefficients of the lag polynomials $a^*(L)$ or $b^*(L)$. The long run effects of shocks can be computed by the following decomposition of the VMA, in obvious notation

$$(6.28)\qquad \begin{pmatrix}r_t\\ x_t\end{pmatrix} = \begin{pmatrix}a^*(1) & b^*(1)\\ c^*(1) & d^*(1)\end{pmatrix}\begin{pmatrix}e_{1t}\\ e_{2t}\end{pmatrix} + (1-L)B^*(L)e_t.$$

The efficient price change is thus equal to the long run component of r_t,

$$(6.29)\qquad w_t = a^*(1)e_{1t} + b^*(1)e_{2t}.$$

The sums of the VMA coefficients, $a^*(1)$ and $b^*(1)$, measure the long run effect of shocks to the system on prices. Assuming that the system is in equilibrium initially, $b^*(1)$ measures the long run price effect of an unexpected shock to the trading process. Thus, the elements of $b^*(1)$ provide an alternative to the non-parametric measures of the adverse selection cost.

---

11 This transformation is described e.g. in Judge et al. (1982, p.657).

6.4.4 The variance of long and short run price effects.

Equation (6.29) shows that there are two sources of changes in the efficient price: shocks to the return equation and shocks to the trade attributes. The former are due to publicly available information that is unrelated to trades, whereas the latter reflect information that is revealed by the trading process. Hasbrouck (1991b) proposes to use the proportion of the variance in the efficient price changes that is due to trading as summary statistic for the informativeness of trades. The variance of w_t is, because e_1 and e_2 are uncorrelated

$$(6.30) \qquad \sigma_w^2 = a^*(1)^2\sigma^2 + b^*(1)\Omega b^*(1)'.$$

The fraction of the variance of w_t that is explained by the attributes of a transaction thus is

$$(6.31) \qquad R^2_{w,x} = \sigma^2_{w,x}/\sigma^2_w, \qquad \sigma^2_{w,x} \equiv b^*(1)\Omega b^*(1)'.$$

The remainder of the variance can be attributed to public information that is unrelated to the trading process.

The transitory deviation of the transaction price from the efficient prices in this model is $s_t=\beta(L)e_t$, where $\beta(L)$ is the first row of $B^*(L)$ in (6.28). Hasbrouck (1993) proposes to use σ_s, the standard deviation of the stationary part of the returns s_t, as a summary measure the quality of a security market; intuitively, σ_s reflects how close the transaction price tracks the efficient price on average. This 'dynamic' measure of transaction costs can be seen as a generalisation of Roll's (1984) estimator. Under Roll's special assumptions, σ_s is equal to half the bid-ask spread (cf. Section 5.7)). The Appendix provides details on the computation of σ_s.

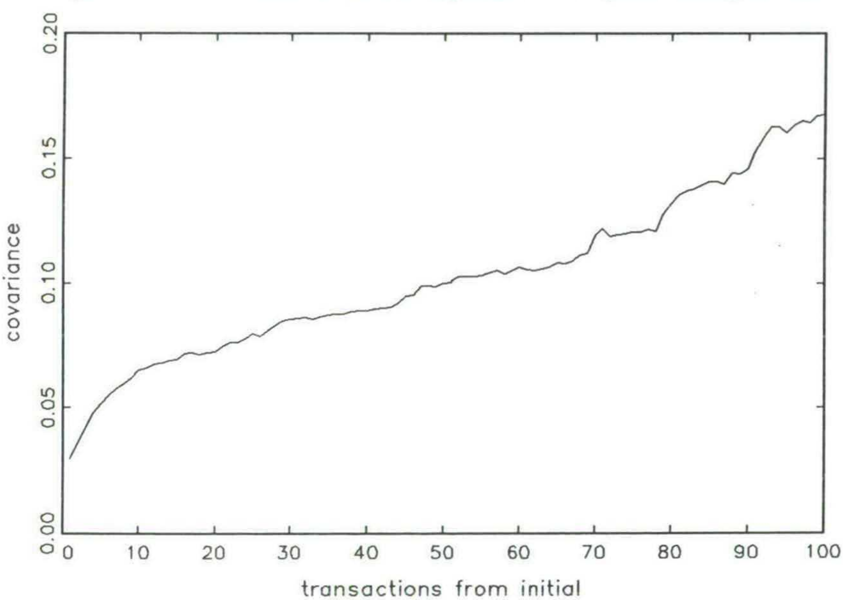
6.5 Empirical results with the simultaneous model.

In this section we report the empirical results of using the simultaneous model of Section 6.4 to estimate the dynamic price effects of trading on the Paris Bourse. Initially, we estimate the price effect of trading by the conditionally expected mid-quote changes given the sign of the transaction only, so that the results can be compared with the results of Section 6.3. Using mid-quotes instead of transaction prices has the advantage that the short run effect is not affected by the processing cost component of the spread, but it relies on the assumption that the mid-point of quotes[12] is a good approximation of the theoretical mid-price y_t.

6.5.1 Non-parametric estimates of price effects.

Figure 6.2 gives an example of the pattern of mid-quote adjustment for intervals between 1 and 100 transactions after the trade, using the non-parametric measure dp_T; the estimates are based on data of the Accor series. The most striking result of this figure is that the initial price effect is much lower than the medium and long run price effect. Even after 100 transactions, there is no price reversal visible. This picture is typical for all series in the sample.

Figure 6.2: Covariance between sign and mid-quote changes Accor



6.5.2 Estimates of the VAR on mid-quote changes and transaction sign.

To facilitate comparison of the results of the VAR model with the non-parametric estimates, we define he returns as the change in the logarithm of the mid-point of the quotes: $r_t=y_{t+1}-y_t$. The VAR is estimated by OLS[13], but because of

[12] The mid-point of the quotes is defined as the average of the best limit buy and sell orders just before the time of the transaction.
[13] Including the sign of the transaction in a simultaneous dynamic model creates

possible heteroskedasticity the standard errors are computed by the method of Newey and West (1987). Overnight returns and opening trades are excluded from the analysis. Following Hasbrouck (1991a), 5 lags of all variables were used as explanatory variables, which seems sufficient given the general absence of residual serial correlation in the estimated equations. Table 6.3 shows an example of an estimated VAR on mid-quote returns and trade sign for the Accor series.

Table 6.3 VAR for return and sign Accor.

	r_t		Q_t	
r_{t-1}	-0.2066	[0.0393]	-1.1408	[0.1723]
r_{t-2}	-0.0719	[0.0189]	-0.5687	[0.1280]
r_{t-3}	-0.0325	[0.0157]	-0.2599	[0.0764]
r_{t-4}	-0.0290	[0.0121]	-0.2246	[0.0789]
r_{t-5}	-0.0227	[0.0120]	-0.1025	[0.0860]
Q_t	0.0291	[0.0008]		
Q_{t-1}	0.0032	[0.0013]	0.2698	[0.0086]
Q_{t-2}	0.0017	[0.0008]	0.1224	[0.0080]
Q_{t-3}	0.0014	[0.0008]	0.0886	[0.0071]
Q_{t-4}	-0.0001	[0.0007]	0.0523	[0.0071]
Q_{t-5}	-0.0003	[0.0008]	0.0801	[0.0070]
s_e	0.0757		0.9063	
r_1	0.0006		-0.0029	

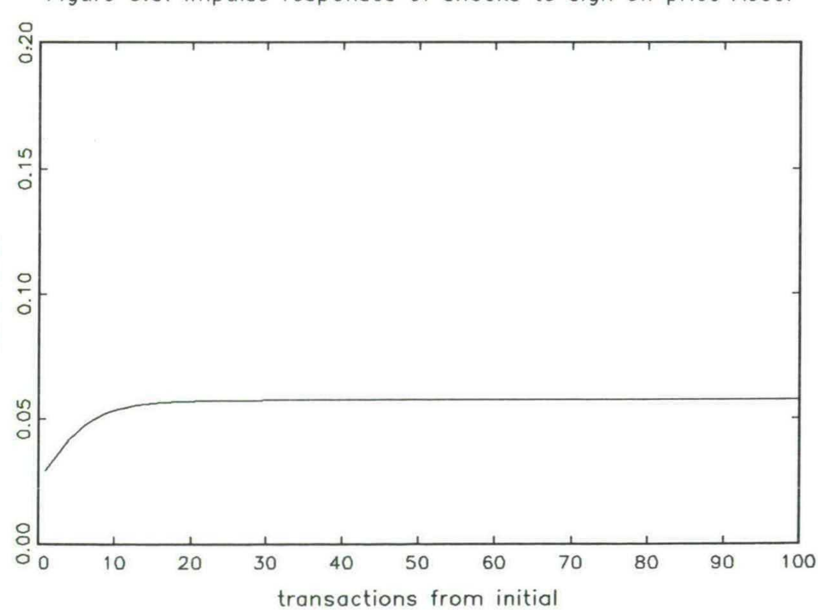
Notes: $r_t = 100\Delta\ln(Y_{t+1})$, Q_t is sign; Y_t is mid-quote before transaction; s_e is standard error of regression; r_1 is serial correlation in errors; Model estimated by OLS, full sample; Newey-West standard errors in brackets.

some problems for estimation and computing dynamic effects, because Q_t is a limited dependent variable that can only take the values -1 and +1. The estimation problem is that (6.23) cannot be a conditional expectation of Q_t for all values of $r_{t-i} \in \mathbb{R}$ if the coefficients of r_{t-i} are non-zero. However, for moderate values of r_{t-i} the linear equation may be a good approximation of the true conditional expectation, and the bias in OLS estimates is probably not too serious. Using Q_t as an explanatory variable in the equation for r_t causes no problems, because the errors of the return equation and the other equations of the VAR are uncorrelated, see Heckman (1978).

The initial effect of a trade on the return can be read immediately from the estimated coefficient of Q_t. For the Accor series, the initial effect is 0.029% of the price, which is almost equal to the 0.027% estimated from the Glosten-Milgrom model (see Table 6.1). There is a strong positive dependence of Q_t on lagged trade sign, which is of course related to the observed positive serial correlation in the Q_t series. The general pattern of sign and magnitude of the estimated VAR coefficients is remarkably similar to the estimates reported by Hasbrouck (1991a,b) for NYSE data.

Figure 6.3 graphs the effects of shocks in the sign on the quote midpoints, obtained by taking partial sums of the VMA coefficients of the returns. The long run effect of a shock in the sign on the price level is estimated at 0.0572%, more than twice the initial effect. The graphs for the other series are very similar to the graph for Accor.

Figure 6.3: impulse responses of shocks to sign on price Accor



6.5.3 Summary of simultaneous analysis of mid-quote changes and sign.

Table 6.4 reports the non-parametric and VAR based estimates of the price effects of a shock to the trade sign for all series under consideration.

Table 6.4 Mid-quote effects of shocks to sign

	dp_1	dp_{100}	VAR_1	VAR_∞
AC	.0294	.1674	.0291	.0576
AQ	.0242	.0707	.0241	.0523
BN	.0181	.0378	.0189	.0284
CA	.0239	.0748	.0239	.0450
CS	.0427	.0836	.0445	.0833
EX	.0168	.0436	.0166	.0317
OR	.0363	.1254	.0374	.0653
RI	.0452	.2437	.0479	.0768
SE	.0436	.1381	.0446	.0835
UAP	.0492	.2287	.0492	.1046

Notes: The table shows the percentage price effects of a transaction by the non-parametric measure in (6.21) and the VAR model (6.26) estimated on mid-quote returns and the transaction sign.
VAR_1 and VAR_∞ are the VAR-based initial and long run price effects estimates.

The results confirm the intuition obtained from Figure 6.1. The initial effect is very small, about 0.02% to 0.05% of the price level. Compare this with the estimated bid-ask spread of 0.15% to 0.35% from Table 6.2. The non-parametric and the VAR-based estimates of the *initial* price effects are very close, but the non-parametric estimate of the *long run* effect is usually much larger than the VAR-based estimate. The non-parametric approximation dp_{100} (the price effect after 100 transactions) ranges from 0.04% to 0.2%, which is much larger than the short run effect. Also the VAR based estimates of the long run price effects are bigger than the short run effects, although smaller than the non-parametric estimates. This confirms the conclusion of the previous section: due to the strong positive serial correlation in the trade sign, the price effects of trading increases over time.

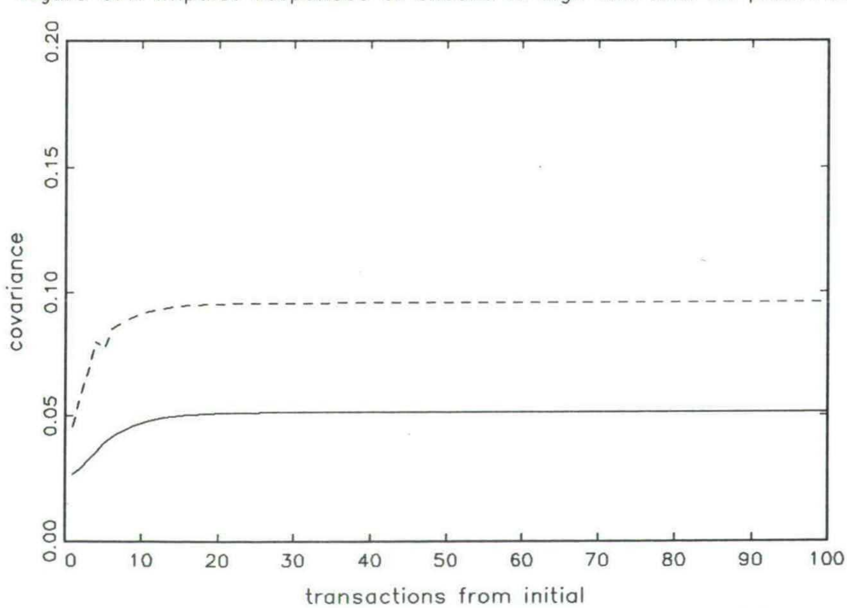
A possible explanation for the difference between non-parametric and VAR based long run price effects is that the VAR estimates are essentially based only on short run covariances between returns and trade attributes. The long run effects in the VAR are computed by extrapolation of the short run dynamics. Campbell and Deaton (1992) demonstrate that small differences in the dynamic specification of a model, which do not change the short run analysis very much, can have a profound influence on the estimates of long run effects.

6.5.4 Simultaneous analysis of mid-quotes, sign and size.

We now extend the VAR model with the transaction size. Five lags of z_t are included as explanatory variables in each equation. For reasons of space, the full estimation results of the trivariate VAR are not reported here, but some specific coefficient estimates deserve some attention. For all series, the current trade size is significant in the equation for the returns, which is not too surprising given that we already know from Section 6.2 that the spread depends on trade size. The sign of the trade shows a very strong positive dependence on lagged trade size. Also in the transaction size there is positive serial correlation, which is in line with the positive serial correlation found in the sign.

The impulse responses to shocks in sign and size follow broadly the same pattern. Typically, both the initial and long run price effect of a transaction of the Normal Market Size is twice the effect of a very small transaction. For both types of shocks, the long run effect is larger than the short run effect. This result is illustrated by Figure 6.4, where the impulse responses of a shock in the sign and a simultaneous shock in sign and size are graphed for the Accor series. These can be interpreted as the dynamic price effects of a very small trade and a trade of the Normal Market Size.

Figure 6.4: impulse responses of shocks to sign and size on price Accor



The short run and long run price effects of a shock in sign and size, computed from the estimated VAR, are summarised in Table 6.5.

Table 6.5 Mid-quote effects of shocks to sign and size

	short run		long run	
	Q_t	z_t	Q_t	z_t
AC	.0262	.0194	.0511	.0422
AQ	.0207	.0175	.0459	.0340
BN	.0162	.0279	.0235	.0551
CA	.0210	.0159	.0785	.0310
CS	.0371	.0488	.0685	.0963
EX	.0147	.0082	.0274	.0149
OR	.0357	.0158	.0564	.0813
RI	.0445	.0258	.0719	.0371
SE	.0421	.0224	.0794	.0372
UAP	.0459	.0192	.0964	.0430

Notes: The table shows the percentage price effects of a transaction by the VAR model (6.29) estimated on transaction price changes, sign and size. The first two columns show the initial price effect of a very small transaction and a transaction of unit size (NMS), respectively. The second pair of columns shows the corresponding long run price effects.

Because the long run effect of a large transaction is larger than the long run effect of a small transaction, we confirm the results of Section 6.2 that the adverse selection cost is positively related to trade size.

6.5.5 Simultaneous analysis of transaction prices, sign and size.

The next step in the analysis is to estimate a VAR on transaction prices rather than mid-quotes, where the returns are now defined as $r_t=p_t-p_{t-1}$. Table 6.6 summarises the most important characteristics of the estimated VAR on transaction prices, sign and size for all ten series. Table 6.6 reports the estimated short run and long run price effects of a transaction.

Note that the initial price effect now estimates all cost components, including the processing cost, which did not appear in the analysis of the mid-quotes. The realised bid-ask spread is estimated as twice the initial price effect. The estimated spreads are very similar to the estimates obtained from the Glosten-Milgrom model in Table 6.1.

Table 6.6 Summary results of VAR on transaction prices.

	Q_t	z_t	S(0)	S(1)	σ_s	σ_w	$R^2_{w,x}$
AC	0.1054 {0.0517}	0.0146 {0.0315}	0.2110 (0.49)	0.2403 (0.69)	0.1038	0.0815	0.40
AQ	0.0843 {0.0444}	-0.0051 {0.0289}	0.1686 (0.52)	0.1584 (0.92)	0.0823	0.0682	0.45
BN	0.0831 {0.0233}	0.0148 {0.0551}	0.1663 (0.28)	0.1960 (0.80)	0.0780	0.0483	0.34
CA	0.0915 {0.0341}	-0.0001 {0.0312}	0.1830 (0.37)	0.1817 (0.72)	0.0931	0.0700	0.28
CS	0.1620 {0.0643}	0.0301 {0.0885}	0.3241 (0.40)	0.3843 (0.79)	0.1662	0.1214	0.35
EX	0.0612 {0.0271}	0.0020 {0.0196}	0.1225 (0.44)	0.1265 (0.74)	0.0620	0.0473	0.38
OR	0.1435 {0.0559}	-0.0001 {0.0799}	0.2869 (0.39)	0.2853 (0.95)	0.1371	0.1013	0.36
RI	0.1548 {0.0710}	-0.0083 {0.0118}	0.3097 (0.46)	0.2830 (0.56)	0.1519	0.1118	0.34
SE	0.1627 {0.0750}	-0.0360 {0.0122}	0.3255 (0.46)	0.2535 (0.69)	0.1514	0.1141	0.36
UAP	0.2071 {0.1074}	-0.0401 {0.0080}	0.4142 (0.52)	0.3324 (0.69)	0.1827	0.1398	0.49

Notes: The first two columns give the leading coefficients of sign and size in the return equation. The sums of the VMA coefficients of sign and size on return (i.e. long run price effects) are reported below the leading coefficients in {}.

The second pair of columns give the estimated bid-ask spread at size 0 and 1 NMS, with the estimated fraction due to adverse selection in () below it.

The fifth column reports the standard deviation of the stationary part of the transaction prices.

The standard deviation of the efficient price changes is reported in the sixth column. The last columns shows the proportion of the variance of the efficient price changes due to the trading process.

The estimates of the long run effects of a trade are similar to the results obtained by the VAR on quote midpoints, only the estimated long run effect of a shock in the trade size is somewhat lower for several firms. This is comforting, because under our assumptions the difference between transaction prices and quote midpoints belongs to the stationary part s_t. Hence, the long run effect of a trade should not be affected by the different definitions of r_t. The long run effect of small transactions is about half of the initial effect, and for NMS the long run effect is often more than 80% of the initial effect. These estimates of the adverse selection cost component are much larger than the estimate obtained from the Glosten-Milgrom model, which assumes that all private information is reflected in the prices immediately after the trade. We already noticed before that this is not the case: the price effect of trading lasts for many more periods.

The estimated standard deviation of the transitory stock price component is nearly equal to half the estimated bid-ask spread. This result is not easily interpreted; if there were no asymmetric information or inventory control this is what you expect. With the large asymmetric information component that we find, σ_s reflects both processing cost and other deviations from the efficient price. There are temporary deviations from the efficient price after trading because it takes some time for the prices to converge to their new equilibrium values. Apparently, the sum of processing costs and temporary deviations from the efficient price is on average about equal to half the bid-ask spread. In short, the 'dynamic' measure (σ_s) proposed by Hasbrouck (1993) yields transaction cost estimates that are not very different from the usual 'static' measure, the realised bid-ask spread.

The final columns of Table 6.6 report the standard deviation of the efficient price changes and the proportion explained by the trading process, $R^2_{w,x}$. In line with the results in Hasbrouck (1991b), we find that between 30% and 40% of the variance of w_t is explained by trading; the remainder is attributable to public information that is unrelated to the trading process. This R^2 has to be interpreted with some caution. Recall that the regression error e_{2t} includes the linearisation error of the discrete sign Q_t, and is therefore a combination of innovations in the trade process and measurement errors. Hence, the variance of e_{2t} is larger than the variance of information revealed by trading. Another word of caution is needed if prices can be quoted only at discrete ticks, or when it is costly to changes quotes or limit orders. In these cases, quoted prices will sometimes deviate from the value predicted by the theory, and only be adjusted after a transaction. If so, the variance of price changes attributed to trades is not completely due to asymmetric information.

6.6 Conclusions.

In this chapter we analysed the intra-day price effects of trading at the Paris Bourse. Special attention was paid to estimating components of the bid-ask spread: processing cost, inventory control cost and adverse selection cost due to asymmetric information. The components were estimated by structural models and by estimates of transitory and permanent price effects of transactions.

We extended the structural model of Madhavan and Smidt (1991) to allow for dependence of inventory control cost and processing cost on the transaction size. It appeared that the inventory control cost cannot be identified if there are no data on the inventory level of market makers available. Therefore, we estimated a simplified model with only processing cost and adverse selection cost. The estimates show that processing cost decreases with transaction size, which explains the finding in Chapter 5 that the bid-ask spread is sometimes decreasing with trade size. In line with theoretical predictions, the adverse selection cost increases with size and accounts for 25% to 40% of the quoted bid-ask spread for average transactions.

The dynamic price effects of trading are analysed both non-parametrically and with a simultaneous model for prices and transaction characteristics such as sign and size. The main result of the dynamic analysis is that the long run price effect of trading is larger than the short run price effect, caused by a strong positive serial correlation in the transaction sign and size. The results sharply contradict the hypothesis that asymmetric information is revealed immediately after a transaction; slow dissemination of information seems more likely. The results are very similar to the findings of Hasbrouck (1991a,b) for NYSE stocks, but contradict those of Holthausen, Leftwich and Mayers (1990), who found that the speed of adjustment to the new equilibrium value after block trades on the NYSE was very fast (less than 3 trades). However, Holthausen et al. analysed the largest transactions in their sample, whereas we censor these transactions. Keim and Madhavan (1992) also present evidence of slow adjustment of prices to new information, although their results are not directly comparable to ours because they analyse only large transactions.

An explanation of the slow price adjustment could be the presence of 'lame duck' limit orders. Many transactions on the Bourse involve one or more limit orders. If prices of limit orders adjust only slowly to new information, probably because of costly monitoring and changing of existing orders, there will from time to time be orders which are either too cheap or too expensive. If so, a number of trades will take place on the same side of the market, causing positive serial correlation in transaction sign and increasing price effects. Berkman (1992) reports similar findings for the European Options Exchange in Amsterdam, which also operates with a public limit order book.

We don't find any evidence for an inventory control effect. At the most general level, inventory control theory predicts a reversal of trade sign and prices, but the empirical results do not show such a pattern at all. The adverse selection cost component is identified by the long run price effects of a transaction. The estimates based on a VAR model for transaction prices suggests that the bid-ask spread can be attributed to adverse selection cost for 50% (for small transactions) up to 80% (for large transactions). These numbers are larger than the estimates obtained from the strucural models, and indeed than many of the numbers reported in the literature (e.g. Stoll (1989)). The reason for this difference is precisely the slow adjustment of prices to new information. If prices adjust slowly, the usual estimators, which are based on the initial and short run effect of a transaction, understate the adverse selection component of the spread.

Appendix. The variance of the stationary part of prices.

The variance of the stationary part of the transaction price, s_t, can be derived as follows. Let $B(L) \equiv \sum_{k=0}^{\infty} b_k L^k$ be the lag polynomial of the VMA (6.28), and let $B^*(L) \equiv (1-L)^{-1}(B(L)-B(1))$. The coefficients of $B^*(L)$ can be derived as follows. First, note that

$$B(L)-B(1) = \sum_{k=0}^{\infty} b_k L^k - \sum_{k=0}^{\infty} b_k = \sum_{k=0}^{\infty} b_k (L^k - 1) = -\sum_{k=0}^{\infty} b_k (1-L)(1+L+ \quad +L^{k-1}).$$

Dividing both sides by (1–L) and rearranging coefficients one obtains

$$B^*(L) = \sum_{k=0}^{\infty} b_k^* L^k, \qquad b_k^* = -\sum_{i=k}^{\infty} b_i.$$

The stationary part of the price is equal to the first element of $B^*(L)e_t$. Therefore, the variance of s_t is equal to the upper-left element of

$$\mathrm{Var}(B^*(L)e_t) = \sum_{k=0}^{\infty} b_k^* \begin{pmatrix} \sigma^2 & 0 \\ 0 & \Omega \end{pmatrix} b_k^{*\prime}$$

All these quantities are readily computed once the lag polynomial of the VAR has been inverted to the polynomial B(L) of the infinite MA representation.

Chapter 7

Summary and Conclusions

In this thesis we analysed some aspects of European financial markets. The first part analysed the impact of the European Monetary System on monetary policy in the EC countries and the behaviour of exchange rates in the EMS target zone. The second part compared the cost of trading French shares on the domestic exchange and the international exchange in London. In this chapter, we briefly summarise the main results of the research and draw some conclusions.

Chapter 2 dealt with the public finance consequences of the EMS membership. In financing public expenditure, the government faces a tradeoff between taxation and seigniorage revenues, i.e. an inflation tax on nominal money holdings. The limited exchange rate flexibility restricts the scope for inflationary policies. Therefore, inflation becomes more costly than without EMS membership. On the other hand, joining a fixed exchange rate regime may increase the credibility of monetary policy because it increases the cost of high inflation. In Chapter 2 it is demonstrated that countries with high costs of tax collection or a high budget deficit may benefit from entering the EMS and hence pegging their currencies to the Deutschemark. If governments follow optimal taxation and seigniorage policies, the model with EMS anchor predicts that tax rates, domestic and foreign inflation are cointegrated. The empirical support for this hypothesis is weak. The model is also tested against a more conventional 'stagflation' model of inflation, and is clearly rejected.

Chapters 3 and 4 analysed EMS exchange rates by target zone models of exchange rate determination. Chapter 3 tested the continuous time Krugman target zone model on EMS data. The basic assumptions of this model are that the target zone is fully credible, defended by marginal interventions only, and that the underlying stochastic process that generates the exchange rate is a Brownian motion. The model is rejected by the data on several points: mean reversion, fat-tailed distributions and conditional heteroskedasticity. The discrete time model of Chapter 4 models these properties explicitly. The empirical results support the discrete time model with a mean-reverting fundamental. This suggests that intra-marginal interventions are an important instrument to stabilise exchange rates in the EMS.

The apparent success of the discrete time specification should not be interpreted as a rejection of continuous time models. There is no a priori economic reason why a discrete time model should be preferred over a continuous time model. The main advantage of discrete time is technical: the discrete time model is easier to solve than the continuous time models for more general stochastic specifications. Another advantage is that standard econometric methods can be used to estimate the discrete time model, whereas simulation techniques or complex likelihood methods are required to estimate continuous time models.

The main deficiency of the models of Chapter 3 and 4 is the assumption that the target zone is fully credible. This hypothesis has strong implications for interest rate differentials. Using the discrete time model to compute expected depreciations within the band, we demonstrate that observed interest rate differentials contain a substantial premium for devaluation risk. The hypothesis that the band is fully credible is clearly rejected for most currencies, even in the most stable period of the EMS (January 1987-September 1992). This point is strongly confirmed by the turmoil in September 1992 and the virtual collapse of the EMS in August 1993. If target zone models are to say anything about a revived EMS or other systems of managed exchange rates, proper modelling of realignment risk seems essential. The discrete time model of Chapter 4 seems a promising starting point for modelling realignments.

Chapter 5 compared the transaction costs on the Paris Bourse and SEAQ International in London. For immediate trades the transaction costs in Paris are small for small transactions, but the costs rise steeply for larger transactions. Often the public limit order book does not provide enough liquidity for even moderately large transactions. The quoted bid-ask spread on the London market is higher than the transaction cost in Paris for small trades. On the other hand, the London market is much deeper: the combined market makers are able to absorb large transactions. An ex-post comparison of the realised spread in Paris and London shows a different picture. The estimated realised spread in Paris is smaller than the estimated realised spread in London for all transaction sizes. One of the causes is that large transactions in Paris are often "crosses" which are negotiated outside the exchange, at a price that is usually within the "fourchette". Another explanation might be that trading in Paris only takes place when the limit order book is relatively deep. Even if we take explicit transaction costs into account, the realised transaction cost in Paris are slightly lower than the cost in London.

The conclusion, therefore, is that the Paris market is nearly always cheaper for small orders, and also for large orders if one is prepared to wait for a counterparty. For an urgent, large transaction the London market is more suited. A point which we did not address in the chapter is that traders might prefer London over Paris because there is no obligation on SEAQ to report transactions to other parties on the market, whereas in Paris all information concerning transactions immediately appears on the public trading screen. Especially traders of large blocks might be afraid that in Paris their order might move the market too much, and prefer to negotiate with market makers in London.

Chapter 6 assesses the price effects of stock transactions on the Paris Bourse. The results show that prices are significantly affected by trading: a buyer initiated transaction pushes subsequent transaction prices upwards, and a seller initiated transaction pushes prices downwards. This effect is caused by new information about the value of the shares that is revealed by trading. The estimates show that the new information is not immediately revealed in prices: price adjustment takes place slowly. A possible explanation for this effect is the presence of 'lame duck' limit orders with too high or too low prices that are not quickly adjusted or withdrawn as soon as new information comes in. An economic rationale for the existence of such orders might be that monitoring or price adjustment is costly.

The analysis can be extended in several directions. We only analysed the majority of small and medium-sized transactions. It would be interesting to estimate the effects of very large transactions. One problem with this is that transaction prices in London, where most large blocks are traded, are not revealed to the public. Another possible extension is to differentiate between buyer- and seller-initiated transactions.

References

Alogoskoufis, George (1990), "Monetary accommodation, exchange rate regimes and inflation persistence", discussion paper No. 390, CEPR, London.

Amemiya, Takeshi (1985), *Advanced Econometrics*, Blackwell, Oxford.

Anderson, Ronald W. and Pierre Tychon (1993), "Competition among European financial markets: The case of Belgian cross-listed stocks", Discussion Paper 9314, Dept. de Sciences Economiques, Université Catholique de Louvain.

Bagehot, Walter (pseud.) (1971), "The only game in town", *Financial Analysts Journal* **22**, 12-14.

Baillie, Richard and Patrick McMahon (1989), *The Foreign Exchange Market: Theory and Econometric Evidence*, Cambridge University Press.

Barro, Robert (1979), "On the determination of the public debt", *Journal of Political Economy* **87**, 940-971.

Barro, Robert (1983), "Inflationary finance under discretion and rules", *Canadian Journal of Economics* **41**, 1-16.

Barro, Robert and David Gordon (1983), "A positive theory of monetary policy in a natural rate model", *Journal of Political Economy* **91**, 589-610.

Bartolini, Leonardo and Gordon Bodnar (1991), "Target zones and forward rates in a model with repeated realignments", mimeo IMF/University of Rochester.

Beetsma, Roel (1991), "Bands and statistical properties of EMS exchange rates", CentER discussion paper 9105, Tilburg University.

Beetsma, Roel (1992), "Imperfect credibility of the band and risk premia in the European Monetary System", CentER discussion paper 9254, Tilburg University.

Beetsma, Roel and Frederick van der Ploeg (1993), "Intramarginal interventions, bands and the pattern of EMS exchange rate distributions", CentER discussion paper 9348, Tilburg University.

Berkman, Henk (1992), *Trading Systems and Liquidity on Securities Markets: A Study of the European Options Exchange*, Ph.D. thesis, Erasmus University Rotterdam.

Bertola, Giuseppe (1993), "Continuous-time models of exchange rates and intervention", in Van der Ploeg, Frederick (ed.), *Handbook of International Economics*, London: Basil Blackwell (forthcoming).

Bertola, Giuseppe and Riccardo Caballero (1992), "Target zones and realignments", *American Economic Review* **82**, 520-536.

Bertola, Giuseppe and Lars Svensson (1993), "Stochastic devaluation risk and the empirical fit of target zone models", *Review of Economic Studies* **60**, 689-712.

Beveridge, S. and C. Nelson (1981), "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the 'business cycle'", *Journal of Monetary Economics* **7**, 151-174.

Biais, Bruno, Pierre Hillion and Chester Spatt (1992), "An empirical analysis of the limit order book and the order flow in the Paris Bourse", paper presented at CEPR workshop on Market Microstructure, 3-4 April 1992, Konstanz.

Breedon, Francis (1992), "Intraday price formation on the London Stock Exchange", working paper Bank of England/LSE.

Bruno, Michael and Jeffrey Sachs (1985). *The Economics of Worldwide Stagflation*, Harvard University Press, Cambridge, MA.

Calvo, Guillermo (1978), "On the time consistency of optimal policy in a monetary economy", *Econometrica* **46**, 1411-1428.

Campbell, John and Angus Deaton (1989), "Why is consumption so smooth?", *Review of Economic Studies* **56**, 357-374.

Canzoneri, Matthew and Carol Rogers (1990), "Is the European Community an optimal currency area? Tax smoothing versus the cost of multiple currencies", *American Economic Review* **80**, 419-433.

Chan, Louis and Josef Lakonishok (1993), "Institutional trades and intraday stock price behavior", *Journal of Financial Economics* **33**, 173-199.

Chernoff, Herman (1954), "On the distribution of the Likelihood Ratio", *Annals of Mathematical Statistics* **25**, 573-578.

Choi, J.Y., Dan Salandro and Kuldeep Shastri (1988), "On the estimation of bid-ask spreads: Theory and evidence", *Journal of Financial and Quantitative Analysis* **23**, 219-230.

Christiano, Larry and Martin Eichenbaum (1989), "Unit roots in real GNP: Do we know, and do we care?", working paper No. 3130, NBER, Cambridge, MA.

Deaton, Angus (1991), "Saving and liquidity constriants", *Econometrica* **59**, 1221-1248.

Deaton, Angus and Guy Laroque (1991a), "On the behaviour of Commodity prices", *Review of Economic Studies* **59**, 1-23.

Deaton, Angus and Guy Laroque (1991b), "Estimating the commodity price model", CREST document de travail no. 9131.

DeJong, Frank, and Frederick van der Ploeg (1991), "Seigniorage, taxes, government debt, and the EMS", CentER Discussion Paper 9155, Tilburg University.

DeJong, Frank, Theo Nijman and Ailsa Röell (1993), "A comparison of the cost of trading French shares on the Paris Bourse and on SEAQ International", CentER Discussion Paper 9329, Tilburg University.

Dixit, Avinash (1992), "The art of smooth pasting", STICERD working paper TE/92/252, London School of Economics.

Domowitz, Ian and Craig Hakkio (1990), "Interpreting an error correction model: Partial adjustment, forward-looking behaviour, and dynamic international money demand", *Journal of Applied Econometrics* **5**, 29-46.

Dornbusch, Rudiger (1988), "The European Monetary System, the dollar and the yen", in Francesco Giavazzi, Stefano Micossi and Marcus Miller, *The European Monetary System*, Cambridge University Press, Cambridge.

Driffill, John, Grayham Mizon and Alastair Ulph (1989), "Costs of inflation", Discussion Paper No. 293, Centre for Economic Policy Research, London.

Duffie, Darrel and Kenneth Singleton (1988), "Simulated moments estimation of diffusion models of asset prices", working paper, Graduate School of Business, Stanford University.

Easley, David and Maureen O'Hara (1987), "Price, trade size and information in securities markets", *Journal of Financial Economics* **19**, 69-90.

Engle, Robert and Clive Granger (1987), "Co-integration and error correction: Representation, estimation and testing", *Econometrica* **55**, 251-276.

Fischer, Stanley and Franco Modigliani (1975), "Towards an understanding of the real effects and costs of inflation", *Weltwirtschaftliches Archiv* **114**, 810-833.

Flood, Robert, Andrew Rose and Donald Mathieson (1990), "An empirical exploration of exchange rate target-zones", NBER working paper 3543.

Friedman, Milton (1969). *The Optimum Quantity of Money and Other Essays*, Aldine Publishing Company, Chicago.

George, Thomas, Gautam Kaul and M. Nimalendran (1991), "Estimation of the bid-ask spread and its components: A new approach", *Review of Financial Studies* **4**, 623-656.

Giavazzi, Francesco and Marco Pagano (1988), "The advantage of tying one's hands: EMS discipline and central bank credibility", *European Economic Review* **32**, 1055-1082.

Giavazzi, Francesco and Alberto Giovannini (1989a), *Limiting Exchange Rate Flexibility: The European Monetary System*, MIT Press, Cambridge, MA.

Giavazzi, Francesco and Alberto Giovannini (1989b), "Monetary interactions under managed exchange rates", *Economica* **56**, 199-214.

Glosten, Lawrence, and Lawrence Harris (1988), "Estimating the components of the bid/ask spread", *Journal of Financial Economics* **21**, 123-142.

Glosten, Lawrence and Paul Milgrom (1985), "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics* **14**, 71-100.

Godfrey, L.G. and M.H. Pesaran (1983), "Tests of non-nested regression models: Small sample adjustments and Monte Carlo evidence", *Journal of Econometrics* **21**, 133-154.

Gourieroux, Christian, Alain Montfort and A. Trognon (1984), "Pseudo Maximum Likelihood methods: Theory", *Econometrica* **52**, 681-700.

Grilli, Vittorio (1989), "Seigniorage in Europe", in: Marcello de Cecco and Alberto Giovannini (eds.), *A European Central Bank? Perspectives on Monetary Unification after Ten Years of the EMS*, Cambridge University Press, Cambridge.

Gros, Daniel (1988), "Seigniorage versus EMS discipline: Some welfare considerations", Working Document No. 38, CEPS, Brussels.

Hansen, Lars Peter (1982), "Large sample properties of Generalized Method of Moments Estimators", *Econometrica* **50**, 1029-1054.

Harris, Lawrence (1986), "A transaction data study of weekly and intra-daily patterns in stock returns", *Journal of Financial Economics* **16**, 99-117.

Harrison, J. Michael (1985), *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.

Hasbrouck, Joel (1988), "Trades, quotes, inventories and information", *Journal of Financial Economics* **22**, 229-252.

Hasbrouck, Joel (1991a), "Measuring the information content of stock trades", *Journal of Finance* **46**, 179-207.

Hasbrouck, Joel (1991b), "The summary informativeness of stock trades: An econometric analysis", *Review of Financial Studies* **4**, 571-595.

Hasbrouck, Joel (1993), "Assessing the quality of a security market: A new approach to transaction cost measurement", *Review of Financial Studies* **6**, 191-212.

Hasbrouck, Joel and George Sofianos (1992), "The trades of market makers: An analysis of NYSE specialists", *mimeo.*

Hausman, Jerry, Andrew Lo and A. Craig MacKinlay (1992), "An ordered probit analysis of transaction stock prices", *Journal of Financial Economics* **31**, 319-379.

Heckman, James J. (1978), "Dummy endogenous variables in a simultaneous equation system", *Econometrica* **46**, 931-959.

Ho, Thomas and Richard Macris (1984), "Dealer bid-ask quotes and transaction prices: An empirical study of some AMEX options", *Journal of Finance* **39**, 23-45.

Ho, Thomas and Hans Stoll (1981), "Optimal dealer pricing under transactions and return uncertainty", *Journal of Financial Economics* **9**, 47-73.

Holthausen, Robert, Richard Leftwich and David Mayers (1990), "Large-block transactions, the speed of response, and temporary and permanent stock-price effects", *Journal of Financial Economics* **26**, 71-95.

Judge, G., R.C. Hill, W. Griffiths, H. Lütkepohl and T. Lee (1982), *Introduction to the Theory and Practice of Econometrics*, Wiley.

Keim, Donald and Ananth Madhavan (1992), "The upstairs market for large-block transactions: Analysis and measurement of price effects", working paper 21-92, Rodney L. White Center for Financial Research, Wharton School, University of Pennsylvania.

Kenen, Peter (1987), "Global policy optimization and the exchange-rate regime", *Journal of Modeling* **9**, 19-63.

Koedijk, Kees, Philip Stork and Casper de Vries (1992), "Between realignments and intervention: The Belgian franc in the European Monetary System", working paper, University of Limburg.

Koedijk, Kees, Philip Stork and Casper de Vries (1993), "An EMS target zone model in discrete time", working paper, Tinbergen Institute Rotterdam.

Kleidon, Allan, and Ingrid Werner (1993), "Round-the-clock trading: Evidence from U.K. cross-listed securities", NBER Working Paper 4410.

Kyle, Pete (1985), "Continuous auctions and insider trading", *Econometrica* **53**, 13-32.

Krugman, Paul (1991), "Target zones and exchange rate dynamics", *Quarterly Journal of Economics* **106**, 669-682.

Krugman, Paul and Marcus Miller (1991), *Exchange Rate Targets and Currency Bands*, Cambridge University Press.

Kydland, Finn E. and Edward C. Prescott (1977), "Rules rather than discretion: The time inconsistency of optimal plans", *Journal of Political Economy* **85**, 473-499.

Lee, Bong-Soo and Beth Fisher Ingram (1991), "Simulation estimation of time-series models", *Journal of Econometrics* **47**, 197-205.

Lindberg, Hans and Paul Söderlind (1992), "Target Zone models and intervention policy: The Swedish case", Institute for International Economic Studies, Stockholm, seminar paper 492.

London Stock Exchange (1992a), "Institutional transaction costs in major European markets", *Stock Exchange Quarterly*, summer edition, 19-26.

London Stock Exchange (1992b), "Price improvement and best execution", *Stock Exchange Quarterly*, spring edition, 25-31.

Mankiw, N. Gregory (1987), "The optimal collection of seignorage: Theory and evidence", *Journal of Monetary Economics* **20**, 327-341.

Madhavan, Ananth (1992), "Trading mechanisms in securities markets", *Journal of Finance* **47**, 607-641.

Madhavan, Ananth and Seymour Smidt (1991), "A Bayesian model of intraday specialist pricing", *Journal of Financial Economics* **30**, 99-134.

Madhavan, Ananth and Seymour Smidt (1992), "An analysis of daily changes in specialist inventories and quotations", *mimeo.*

McInish, Thomas and Robert Wood (1990), "An analysis of transactions data for the Toronto stock exchange", *Journal of Banking and Finance* **14**, 441-458.

McFadden, Daniel (1989), "A method of simulated moments for estimation of discrete response models without numerical integration", *Econometrica* **57**, 995-1026.

McKibbin, Warwick and Jeffrey Sachs (1988), "Comparing the global performance of alternative exchange arrangements", *Journal of International Money and Finance* **7**, 387-410.

Meese, Richard and Kenneth Rogoff (1983), "Empirical exchange rate models of the seventies: Do they fit out-of-sample?", *Journal of International Economics* **14**, 3-24.

Miller, Marcus and Paul Weller (1991), "Exchange rate bands with price inertia", *Economic Journal* **101**, 1380-1399.

Mussa, Michael (1986), "Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications", in K. Brunner and A. Meltzer (eds.), *Carnegie-Rochester Conference Series on Public Policy*, **25**, 117-124.

Newey, Whitney (1985), "Maximum Likelihood specification testing and conditional moment tests", *Econometrica* **53**, 1047-1069.

Newey, Whitney and Kenneth West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* **55**, 703-708.

Niemeyer, Jonas and Patrick Sandas (1992), "The market micro structure of the Stockholm stock exchange and the Swedish options market", mimeo.

Obstfeld, Maurice (1991), "Dynamic seigniorage theory: An exploration", *Journal of Political Economy*, forthcoming.

Pagan, Adrian and Michael Wickens (1989), "A survey of some recent econometric methods", *Economic Journal* **99**, 962-1025.

Pagano, Marco and Ailsa Röell (1992), "Auction and dealership markets: What's the difference?", *European Economic Review* **36**, 613-623.

Pesaran, M. Hashem and Hossein Samiei (1992a), "An analysis of the determination of the DM/FF exchange rate in a discrete-time target-zone model", *Economic Journal* **102**, 388-401.

Pesaran, M. Hashem and Hossein Samiei (1992b), "Limited dependent rational expectations models with future expectations", working paper, Cambridge University.

Phelps, Edward (1973), "Inflation in the theory of public finance", *Swedish Journal of Economics* **75**, 67-82.

Phillips, Peter (1987), "Time series regression with a unit root", *Econometrica* **55**, 277-301.

Ploeg, Frederick van der (1989), "Monetary interdependence under alternative exchange-rate regimes: A European perspective", Discussion Paper 358, Centre for Economic Policy Research, London.

Ploeg, Frederick van der (1991), "Unanticipated inflation and government finance: The case for an independent common central bank", mimeo, CentER, Tilburg University.

Poterba, Jim and Julio Rotemberg (1990), "Inflation and taxation with optimizing governments", *Journal of Money, Credit and Banking* **22**, 1-18.

Röell, Ailsa (1992), "Comparing the performance of stock exchange trading systems", in Fingleton and Schoenmaker (eds.), *The Internationalisation of Capital Markets and the Regulatory Response*, Graham and Trotman, London.

Roll, Richard (1984), "A simple implicit measure of the effective bid-ask spread in an efficient market", *Journal of Finance* **39**, 1127-1139.

Rose, Andrew and Lars Svensson (1991), "Expected and predicted realignments: The FF/DM exchange rate during the EMS", NBER working paper 3695 and IIES seminar paper 485.

Sargent, Thomas (1979). *Macroeconomic Theory*, Academic Press, New York.

Sargent, Thomas (1987), *Dynamic Macroeconomic Theory*, MIT Press, Cambridge, MA.

Schmidt, Hartmut and Peter Iversen (1991), "Geld-Brief-Spannen deutscher Standartwerten in IBIS und MATIS", *Zeitschrift für Bankrecht und Bankwirtschaft* **3**, 209-226.

Smith, Gregor and Michael Spencer (1991), "Estimation and testing in models of exchange rate target zones and process switching", in P. Krugman and M. Miller (eds.), *Exchange Rate Targets and Currency Bands*, Cambridge University Press.

Stoll, Hans (1989), "Inferring the components of the bid-ask spread: Theory and empirical tests", *Journal of Finance* **44**, 115-134.

Svensson, Lars (1991a), "The simplest test of target zone credibility", *IMF staff paper* **38**, 655-665.

Svensson, Lars (1991b), "The term structure of interest rate differentials in a target zone: Theory and Swedish data", *Journal of Monetary Economics* **28**, 87-116.

Svensson, Lars (1991c), "Target zones and interest rate variability", *Journal of International Economics* **31**, 27-54.

Svensson, Lars (1991d), "Assessing target zone credibility: Mean reversion and devaluation expectations in the EMS", NBER working paper 3795, forthcoming in *European Economic Review*.

Svensson, Lars (1992), "An Interpretation of Recent Research on Exchange Rate Target Zones", *Journal of Economic Perspectives* **6**, 119-144.

Taylor, John (1979), "Staggered wage setting in a macro model", *American Economic Review*, Papers and Proceedings **69**, 108-113.

Tonks, Ian and Andy Snell (1992), "Trading volumes and stock market prices", LSE Financial Markets Group Discussion Paper 130.

Trehan, B. and C. Walsh (1990), "Seigniorage and tax smoothing in the United States, 1914-1986", *Journal of Monetary Economics* **25**, 97-112.

Turnovsky, Stephen J. and William A. Brock (1980), "Time consistency and optimal government policies in a perfect foresight system", *Journal of Public Economics* **13**, 183-212.

Van Wijnbergen, Sweder (1991), "Fiscal deficits, exchange rate crises and inflation", *Review of Economic Studies* **58**, 81-92.

Wickens, Michael and Trevor S. Breusch (1988), "Dynamic specification, the long-run and the estimation of transformed regression models", *Economic Journal* **98**, S189-S205.

Worthington, Peter (1991), "Global equity turnover: Market comparisons", *Bank of England Quarterly Review*, May, 246-249.

Samenvatting

Dit proefschrift bestaat uit twee delen. Het eerste deel heeft als onderzoeksobject het Europese Monetaire Stelsel (EMS) en het tweede deel de microstructuur van aandelenmarkten, met name de markt in Franse aandelen. **Hoofdstuk 1** geeft een inleiding tot beide onderwerpen.

De belangrijkste factor in het monetaire beleid van de laatste 15 jaar in Europa is het EMS, een stelsel van vaste, maar aanpasbare wisselkoersen tussen de valuta van de belangrijkste landen in de EEG. In het EMS kunnen de wisselkoersen beperkt fluctueren in een band rond de spilkoers. Om de wisselkoers van hun munt binnen de gegeven bandbreedtes te houden intervenieren de centrale banken in tijden van spanning in de valutamarkt door middel van de aan- en verkoop van grote partijen buitenlandse valuta. Aanpassingen van de spilkoers kan alleen plaatsvinden met toestemming van de andere deelnemers aan het EMS.

Hoofdstuk 2 beschouwt de gevolgen van deelname aan het EMS voor de inkomsten van de overheid uit geldschepping, hetgeen voor met name zuid-Europese landen traditioneel een belangrijke bron van inkomsten is. Deelname aan het EMS dwingt landen tot stabilisatie van hun wisselkoers en dus tot lagere inflatie en een krapper geldbeleid. Om toch de overheidsuitgaven te financieren zullen andere belastingen, zeker op lange termijn, omhoog moeten. Als het innen van belastingen kostbaar is, lijkt dit een ongunstig effect van toetreding tot het EMS. Anderzijds kan het EMS juist ook gebruikt worden om een geloofwaardig anti-inflatiebeleid van de grond te krijgen. Voor landen met een grote staatsschuld of een inefficiënte belastinginning kan toetreding tot het EMS geloofwaardigheid van monetair beleid 'kopen'. In hoofdstuk 2 wordt ook enig empirisch werk naar de afweging tussen monetaire financiering en belastingheffing gerapporteerd. De theorie voorspelt dat belastingvoeten, inflatie en de depreciatie van de wisselkoers gecoïntegreerd zijn. Veel ondersteuning voor de theorie vinden we niet, behalve voor enige landen met een traditioneel hoge inflatie. De theorie wordt geconfronteerd met een alternatief 'stagflatie' model, waarin inflatie wordt verklaart uit prijsschokken van grondstoffen (olie) en een loon-prijsspiraal. Dit model blijkt het empirisch veel beter te doen dan het eerste model.

Naast deze macro-economische implicaties van het EMS is er een veel directere invloed van het EMS op de wisselkoersen zelf en op de rentestanden. Het ligt voor de hand dat door de beperkte fluctuatiemarges de wisselkoersen stabieler zijn dan in een stelsel van volledig vrije wisselkoersen. Krugman (1991) heeft een model onwikkeld ter beschrijving van wisselkoersen binnen banden. De achterliggende gedachte van het model is dat de wisselkoers een speculatieve prijs is die tot stand komt als de verdisconteerde waarde van economische 'fundamentals', met name de relatieve geldhoeveelheid in de betrokken landen. De geldhoeveelheid is in een vast wisselkoersregime echter een gereguleerde grootheid. De monetaire overheid beïnvloedt de geldhoeveelheid onder andere door interventies in de valutamarkt. Rationele speculanten nemen de mogelijkheid van interventies mee in hun bepaling van de wisselkoers. Krugman toont aan dat hierdoor de wisselkoers ook binnen de band gestabiliseerd wordt.

Hoofdstuk 3 onderzoekt de geldigheid van het model van Krugman voor EMS wisselkoersen. In het hoofdstuk wordt een nieuwe schattingsprocedure voor de parameters van dit model voorgesteld. Aangezien het model in continue tijd is gespecificeerd, is er een vertaalslag van het continue model naar een conditionele verdeling voor de op discrete tijdstippen waargenomen wisselkoersen nodig. Met behulp van deze verdeling construeren we een schatter die een zeer goede benadering is van de maximale aannemelijkheidsschatter. De voorgestelde schatter is efficiënter dan alternatieven die gebaseerd zijn op simulatie van het model. De empirische resultaten zijn helaas niet erg gunstig: weliswaar wordt er enige ondersteuning voor stabilisatie van wisselkoersen binnen de band gevonden, maar de specificatie van het model laat te wensen over. Relatief grote wisselkoersveranderingen blijken vaker voor te komen dan het model voorspelt en bovendien varieert de volatiliteit van de koersen meer door de tijd dan door het model voorspeld wordt.

Hoofdstuk 4 probeert een aantal van de met Krugmans model gesignaleerde problemen weg te nemen. Hiertoe specificeren we het model direct in discrete tijd in plaats van continue tijd. Dit heeft een aantal voordelen. Ten eerste is het model in discrete tijd gemakkelijker en onder algemenere voorwaarden op te lossen dan het model in continue tijd. Ten tweede zijn de statistische eigenschappen van het model gemakkelijker te analyseren en de zijn benodigde schatters eenvoudiger. Het discrete tijd model, uitgebreid met een dikstaartige verdeling van de 'fundamental' en interventies binnen de band, blijkt aanzienlijk beter in staat de waargenomen wisselkoersen te beschrijven dan de veel beperktere specificatie van Krugman. Een nadeel van de analyse blijft echter dat de band volledig betrouwbaar wordt verondersteld. In hoofdstuk 4 wordt daarom tevens onderzocht hoe betrouwbaar de band is. Aan de hand van waargenomen renteverschillen, die gecorrigeerd worden voor de verwachte depreciatie binnen de band, wordt aangetoond dat de EMS banden nooit volledig betrouwbaar werden geacht, al is de geloofwaardigheid na verloop van tijd wel toegenomen. Een open einde is het modelleren van herschikkingen van de spilkoersen. De aanpak in discrete tijd, zoals voorgesteld in Hoofdstuk 4, lijkt hiervoor veelbelovend.

Het tweede deel van het proefschrift behandelt de microstructuur van aandelenmarkten, met name de markten in Franse aandelen. De twee belangrijkste markten voor Franse aandelen zijn de Bourse in Parijs en de Franse sector van SEAQ International in Londen. Deze handel op deze markten loopt volgens geheel verschillende systemen. Ons onderzoek heeft mede tot doel om de implicaties van dit verschil voor de kwaliteit van deze beurzen te meten.

Hoofdstuk 5 vergelijkt de transactiekosten op de Bourse en SEAQ International. Zowel voor een (hypothetische) urgente transactie als voor de feitelijk gerealiseerde transacties wordt het gemiddelde bied-laat koersverschil (de 'spread') geschat. Hiertoe gebruiken we zowel niet-parametrische schatters als een lineair model, dat minder gevoelig is voor een aantal onvolkomenheden in de data. De resultaten laten zien dat de impliciete transactiekosten voor een urgente transactie in Parijs snel oplopen voor grotere hoeveelheden. In Londen daarentegen is de spread voor kleine transacties relatief groot, maar voor grote transacties juist vrij gunstig. De spread voor gerealiseerde transacties is in Parijs kleiner dan in Londen voor elke transactieomvang. Als we de expliciete transactiekosten meenemen blijkt Parijs goedkoper voor kleine transacties, en ongeveer even duur als Londen voor grote transacties. Het grote voordeel van de Londense markt is echter de mogelijkheid om tegen relatief gunstige tarieven grote orders snel te verwerken.

Hoofdstuk 6 beschouwt de prijseffecten van een transactie op de Parijse Bourse. Met name zijn we geïnteresseerd in het lange termijn effect, dat iets zegt over de nieuwe informatie die door de transactie op de markt is gekomen. Het doel van de analyse is om de kostencomponenten van de spread te meten. De literatuur onderscheidt drie kosten: kosten van orderverwerking, voorraadbeheerskosten en kosten die voortvloeien uit adverse selectie als gevolg van asymmetrische informatie. Uit de schattingen blijkt dat de eerste en de derde component significant zijn, en bovendien van de grootte van de transactie afhangen: de kosten van orderverwerking (per aandeel) nemen af voor grotere transacties, terwijl de adverse selectie component van de spread juist toeneemt met de transactieomvang. Er wordt geen bewijs gevonden voor het bestaan van voorraadbeheerskosten. Een ander resultaat is dat nieuwe informatie langzaam in de koersen verwerkt wordt. Zelfs na enige tientallen transacties is er nog een merkbaar prijseffect.